

STIC-Biotech/ChemLib

From: Jiang, Dong  
Sent: Thursday, May 15, 2003 6:49 PM  
To: STIC-Biotech/ChemLib  
Subject: 10/050,704

94181

RECEIVED  
Priority: 4/9/99  
MAY 16 2003  
HGS (Soppet, D.)  
(STIC)

Please search SEQ ID NO:125

-issued & Pub.  
-commercial

no utility  
Point of Contact:  
Toby Port  
Technical Info. Specialist  
CM1 6A64  
703-303-3534  
# - 1/18/02

Please send results on paper to Dong Jiang in 10D-08 (mail stop CM1-10D19).  
Thank you very much.

Dong Jiang (78243)  
703-305-1345  
U.S. Patent and Trademark Office  
Art Unit 1646  
dong.jiang@uspto.gov  
CM1-10D08  
Mail stop: CM1-10D19

Do a alignment b/w SEQ 125 & 127.

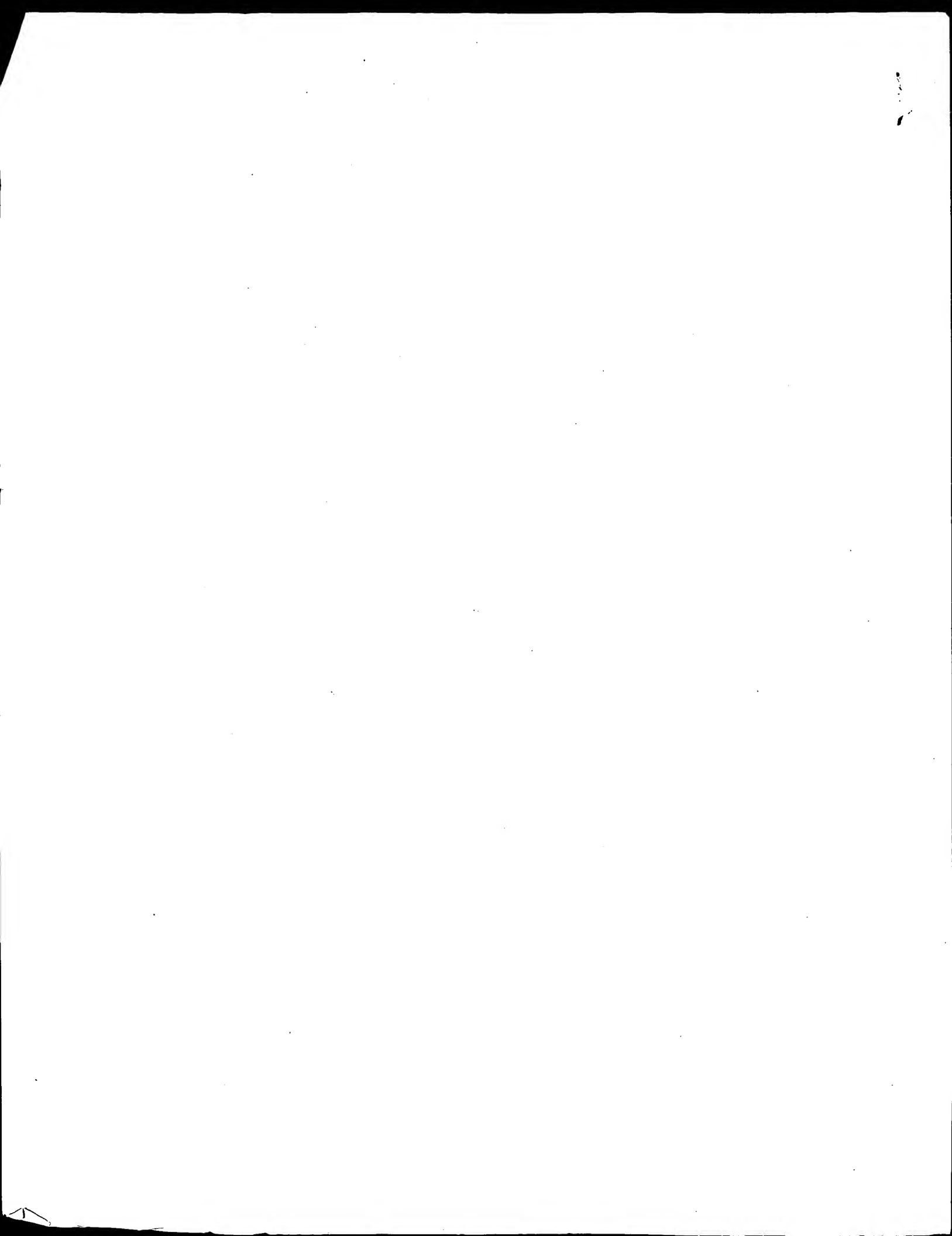
Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 5/16  
Date Completed: 5/16  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 10:45:57 ; Search time 40 Seconds  
(without alignments)  
2088.703 Million cell updates/sec

Title: US-10-050-704-125

Perfect score: 3208

Sequence: 1 MEARVHALKRGVSLLCVF.....KATEPLGFWAHNGEVRIA 627

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.GeneSeq.101002.\*  
1: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3208	100.0	627	21	Human secreted pro
2	2766	86.2	889	22	Novel human diago
3	2587.5	80.7	834	22	Novel human diago
4	2473.5	77.1	653	22	Novel human diago
5	2304	71.8	797	22	Novel human diago
6	2287	71.3	787	17	Novel human diago
7	2249	70.1	800	21	Human protocadheri
8	2249	70.1	800	22	Human protocadheri
9	2237	69.7	1888	22	Human PRO polypept
10	2142	66.8	798	20	Human PRO531 prote

11	2142	66.8	798	21	AA844302
12	2142	66.8	798	21	AA824049
13	2142	66.8	798	21	AA833433
14	2142	66.8	798	23	AA83657
15	2118	66.0	779	21	AA94997
16	2009.5	62.6	797	17	AA86866
17	1879	58.6	744	22	AA89498
18	1841	57.4	743	21	AA89492
19	1576.5	49.1	552	22	AA89487
20	1558	48.6	544	23	ABP42004
21	1457	45.4	559	23	AB872287
22	1418	44.2	931	22	AA878649
23	1418	44.2	949	22	AB812315
24	1418	44.2	949	22	AA879633
25	1401	43.7	833	22	AB826040
26	1400	43.6	858	22	AA879446
27	1393	43.4	931	22	AA878466
28	1383.5	43.1	932	22	AB817156
29	1377.5	42.9	832	22	AB823820
30	1343.5	41.9	916	20	AA841743
31	1343.5	41.9	916	21	AA844399
32	1343.5	41.9	916	22	AA829062
33	1323	41.2	820	22	AB823824
34	1288	40.1	1958	22	AB821821
35	1257	39.5	459	22	AA892859
36	1257	39.2	405	22	AA893330
37	1232.5	38.4	934	23	AA847925
38	1217.5	38.0	1008	22	AB815334
39	1217	37.9	682	15	AA849144
40	1217	37.9	682	17	AA87154
41	1217	37.9	836	15	AA858912
42	1217	37.9	836	17	AA87153
43	1217	37.9	904	15	AA858907
44	1217	37.9	904	17	AA87147
45	1192.5	37.2	949	22	AB823232

#### ALIGNMENTS

RESULT 1	AA838349	standard; Protein; 627 AA.
ID	AA838349	
XX	AA838349	
AC	AA838349	
XX	AA838349	
DT	31-JAN-2001	(first entry)
DE	Human secreted protein encoded by gene 29 clone HOFND85.	
XX	Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;	
XX	cytostatic; cardiant; vasotropic; cerebroprotective; neuroprotective;	
KW	neotropic; antibacterial; virucide; fungicide; ophthalmological; human;	
KW	vulnary; gene therapy; infection; secreted protein.	
OS	Homo sapiens.	
XX	MO200061623-A1.	
PN	19-OCT-2000.	
XX	06-APR-2000; 2000WO-US08979.	
PF	09-APR-1999; 99US-0128693.	
PR	26-APR-1999; 99US-0130991.	
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	Ruben SM, Ni J, Komatsu G, Soppet DR, Shi Y,	
PI	Latteur DM, Olsen HS, Ebner R, Florence KA, Moore PA, Blitze CE,	
PI	Young PE;	
DR	WPI: 2000-647418/62.	

XX New nucleic acid molecules encoding 62 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -  
 XX  
 PS Claim 11; Page 619-621; 716pp; English.  
 XX  
 CC Sequences AAB38321-B38396 represent the amino acid sequences of 62  
 CC human secreted proteins encoded by the genes AAG69512-C69587. The genes  
 CC and proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
 CC a range of human tissues disclosed in the specification. The nucleic  
 CC acids, proteins, antibodies and (ant)agonists are useful in the  
 CC diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.  
 CC rheumatoid arthritis; (b) hyperproliferative disorders e.g. neoplasms  
 CC of the breast or liver; (c) cardiovascular disorders e.g. cardiac  
 CC arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)  
 CC angiogenesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)  
 CC infections caused by bacteria, viruses and fungi; and (h) ocular  
 CC disorders e.g. corneal infection. The polypeptides can also be used to  
 CC aid wound healing and epithelial cell proliferation, to prevent skin  
 CC aging due to sunburn, to maintain organs before transplantation, for  
 CC supporting cell culture of primary tissues, to regenerate tissues and in  
 CC chemotaxis.

XX Sequence 627 AA:

Query Match 100.0%; Score 3208; DB 21; Length 627;  
 Best Local Similarity 100.0%; Pred. No. 6,5e-278;  
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEARVHALOKROYSLVCLVFLVSMAGAEPLRFVAEETGFFFLANLAIDGLGVEELS 60  
 DB 1 MEARVHALOKROYSLVCLVFLVSMAGAEPLRFVAEETGFFFLANLAIDGLGVEELS 60  
 QY 61 ARGRIVSDEITGFLINPLTGLLNENKIDREELGPEPCVLPOLLLEKPEQFRAE 120  
 DB 61 ARGRIVSDEITGFLINPLTGLLNENKIDREELGPEPCVLPOLLLEKPEQFRAE 120  
 QY 121 LWRVINDHSPVFLDREITINLESTPGATFLLESADSDVGINLRNVTISSNVYFHI 180  
 DB 121 LWRVINDHSPVFLDREITINLESTPGATFLLESADSDVGINLRNVTISSNVYFHI 180  
 QY 181 NVHDSNGVNYSELVLDKVDREVEVELRLTGLDGSPPSSGTLITILVDINDVP 240  
 DB 181 NVHDSNGVNYSELVLDKVDREVEVELRLTGLDGSPPSSGTLITILVDINDVP 240  
 QY 241 EFVESLYKQVPEPNSPVGLVYVSARDLDTGSGNCEIVAFVATERTLKTIRINSTGN 300  
 DB 241 EFVESLYKQVPEPNSPVGLVYVSARDLDTGSGNCEIVAFVATERTLKTIRINSTGN 300  
 QY 301 LHKALNEAIOITYTLTQAKDGGGLSKCTVVVHVTDINDPPELMSLTSPSPENS 360  
 DB 301 LHKALNEAIOITYTLTQAKDGGGLSKCTVVVHVTDINDPPELMSLTSPSPENS 360  
 QY 361 PETVVAVFRIRDRSGNNKMYCSIODHLPEVLKSVSEVFYTLVERALDREEREYNT 420  
 DB 361 PETVVAVFRIRDRSGNNKMYCSIODHLPEVLKSVSEVFYTLVERALDREEREYNT 420  
 QY 421 IVYTDIGTRELKTOHNLVTVSADVNDNAFTSQTTYTLKRENNSPALHIGSVATDSDS 480  
 DB 421 IVYTDIGTRELKTOHNLVTVSADVNDNAFTSQTTYTLKRENNSPALHIGSVATDSDS 480  
 QY 481 GANAQVYSLRPHDPQLPLGSLVSIINDNGOLFALRSIDFALQAFERRGADRGSPA 540  
 DB 481 GANAQVYSLRPHDPQLPLGSLVSIINDNGOLFALRSIDFALQAFERRGADRGSPA 540  
 QY 541 LSSQALVRLVADANDNAFVLYPLONGSAPCTELVPRAAEGYVAVVAVDDGSGONA 600  
 DB 541 LSSQALVRLVADANDNAFVLYPLONGSAPCTELVPRAAEGYVAVVAVDDGSGONA 600  
 QY 601 WLSTYOLKATPEGLFGVMAHNGEVRTA 627  
 DB 601 WLSTYOLKATPEGLFGVMAHNGEVRTA 627

DB 601 WLSTYOLKATPEGLFGVMAHNGEVRTA 627

RESULT 2

ID ABG23879  
 ABG23879 standard; Protein; 889 AA.

XX ABG23879;

XX 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #23870.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Dymnac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS88066.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX Claim 20; SEQ ID NO 54238; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 889 AA:

Query Match 86.2%; Score 2766; DB 22; Length 889;  
 Best Local Similarity 85.8%; Pred. No. 4.2e-238;

Matches 538; Conservative 32; Mismatches 57; Indels 0; Gaps 0;

QY 1 MEARVHALOKROYSLVCLVFLVSMAGAEPLRFVAEETGFFFLANLAIDGLGVEELS 60  
 DB 95 MEARVERAVQKQVFLVCLVFLVSMAGAEPLRFVAEETGFFFLANLAIDGLGVEELS 154



QY 61 ARGCRIVSDETIGELLNPLTGDLINKRELCGTEPCVLPFCULLKEKPOIFRAE 120  
 DB 155 ARGTRIVSDQNMQILLSSLTGDLINKRELCGTEPCVLPFCULLKEKPOIFRAE 214  
 QY 121 LNVARDINHSVPFLDREITLNTLESTPGATFLLESANDSDVGINLNNTYISSNVEYFI 180  
 DB 215 LNVARDINHSVPFLDREITLNTLESTPGATFLLESANDSDVGINLNNTYISSNVEYFI 274  
 QY 181 NVHNDGEGNVSELYLDKYLDEEVEPELRLTLTGDDGSPRSGTTLRLIYLDINDVNP 240  
 DB 275 NVHDSGEGNIYELVYNQVLDREELPEPSLTTLTALDGSPPRSGTALVRIYLDVNDAP 334  
 QY 241 EYESLKYQVPENSPVGSVYTVSARDIDTSGNEIYAFYATERTLKEFRINSTSGN 300  
 DB 335 DEVRSLKYQVPENSPVGSVYTVSARDIDTSGNEIYAFYATERTLKEFRINSTSGN 394  
 QY 301 LHLKAELEAYEAIQTYTLTQAKDGGLSGCTVYVAVTDINDNPELLMSLTSPFENS 360  
 DB 395 LHLKAELEAYEAIQTYTLTQAKDGGLSGCTVYVAVTDINDNPELLMSLTSPFENS 454  
 QY 361 PETVAVAFRIIRDSDGNNAKWCSTIODHLPEYLKDSVENFTYLERALDREEREYNT 420  
 DB 455 PETVAVAFRIIRDSDGNNAKWCSTIODHLPEYLKDSVENFTYLERALDREEREYNT 514  
 QY 421 ITVTDGTRPKTOHNLTVTVSDVNDNAPTFESQTYTLTVRENNSEFALHIGSVATDRDS 480  
 DB 515 ITVTDGTRPKTOHNLTVTVSDVNDNAPTFESQTYTLTVRENNSEFALHIGSVATDRDS 574  
 QY 481 GANAQVYISLPSQDPLPLASVSTINADNGOLFALRSIDFEALQAFERRVGAADGSPA 540  
 DB 575 GANAQVYISLPSQDPLPLASVSTINADNGOLFALRSIDFEALQAFERRVGAADGSPA 634  
 QY 541 LSSQALVRYLVANANAPFVLPYLDNGSAPCELPRAAEAGYLAKVVAVDGDSQNA 600  
 DB 635 LSSQALVRYLVANANAPFVLPYLDNGSAPCELPRAAEAGYLAKVVAVDGDSQNA 694  
 QY 601 WLSTYQLLKATEPGLFGVMAHNGEVRPA 627  
 DB 695 WLSTYQLLKATEPGLFGVMAHNGEVRPA 721  
 RESULT 3  
 ABG23876 ID ABG23876 standard; Protein: 834 AA.  
 AC ABG23876:  
 DT 18-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #23867.  
 DE Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 OS WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001, 2001WO-US08631.  
 PF 31-MAR-2000, 2000US-0540217.  
 PR 23-AUG-2000, 2000US-0649167.  
 XX (HSE-) HYSED INC.  
 PA Dimaenac RT, Liu C, Tang YT;  
 PI WPI: 2001-639362/73.  
 DR N-PSDB: AAS88063.  
 XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX Claim 20: SEQ ID No 54235; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SO Sequence 834 AA;  
 Query Match 80.7%; Score 2587.5; DB 22; Length 834;  
 Best Local Similarity 74.9%; Pred. No. 3.6e-22;  
 Matches 517; Conservative 37; Mismatches 61; Indels 75; Gaps 3;  
 QY 1 MEARVHALOKROVSLCLVFVGVSWAGAEPLRYFAEETERTGLANLAIDGLVEELS 60  
 DB 1 MEARVRAVOKROYFLVFLVGLMSWAGAEPLRYFAEETERTGLNANLGLGVGELR 60  
 QY 61 ARGCRIVSDETIGELLNPLTGDLINKRELCGTEPCVLPFCULLKEKPOIFRAE 120  
 DB 61 ARGTRIVSDQNMQILLSSLTGDLINKRELCGTEPCVLPFCULLKEKPOIFRAE 120  
 QY 121 LNVARDINHSVPFLDREITLNTLESTPGATFLLESANDSDVGINLNNTYISSNVEYFI 180  
 DB 121 LNVARDINHSVPFLDREITLNTLESTPGATFLLESANDSDVGINLNNTYISSNVEYFI 180  
 QY 181 NVHNDGEGNVSELYLDKYLDEEVEPELRLTLTGDDGSPRSGTTLRLIYLDINDVNP 240  
 DB 181 NVHDSGEGNIYELVYNQVLDREELPEPSLTTLTALDGSPPRSGTALVRIYLDVNDAP 240  
 QY 241 EYESLKYQVPENSPVGSVYTVSARDIDTSGNEIYAFYATERTLKEFRINSTSGN 300  
 DB 241 DEVRSLKYQVPENSPVGSVYTVSARDIDTSGNEIYAFYATERTLKEFRINSTSGN 300  
 QY 301 LHLKAELEAYEAIQTYTLTQAKDGGLSGCTVYVAVTDINDNPELLMSLTSPFENS 360  
 DB 301 LHLKAELEAYEAIQTYTLTQAKDGGLSGCTVYVAVTDINDNPELLMSLTSPFENS 360  
 QY 361 PETVAVAFRIIRDSDGNNAKWCSTIODHLPEYLKDSVENFTYLERALDREEREYNT 420  
 DB 361 PETVAVAFRIIRDSDGNNAKWCSTIODHLPEYLKDSVENFTYLERALDREEREYNT 420  
 QY 421 ITVTDGTRPKTOHNLTVTVSDVNDNAPTFESQTYTLTVRENNSEFALHIGSVATDRDS 480  
 DB 421 ITVTDGTRPKTOHNLTVTVSDVNDNAPTFESQTYTLTVRENNSEFALHIGSVATDRDS 480  
 QY 481 GANA-----  
 DB 481 GANA-----  
 QY 485 -----QVTSLLPRLPHDPQLPLGLSVSTINADNGOLFALRSIDFEALQAFERRVGAAD 536  
 DB 541 VSTAETNIQVTVSLPSQDPLPLASVSTINADNGOLFALRSIDFEALQAFERRVGAAD 600

OY 537 GSPALSSQALVRLVADANDNAPFLVPLQNGSAPCTELVPR-AAEAGYLVAKVAVDGD 595  
 DB 601 GSTALSSSEALVRLVLDANDSSLFVLPPLQNGSAPCTELMPRAAERGGLVTK----- 653  
 OY 596 SGNAMLSYQLLKATEPGLFGVMAHNGEVR 625  
 DB 654 ---NMLSLXQLLKATEPGLFGVMAHNGTDR 680

## RESULT 4

ABG23877 ID ABG23877 standard; Protein; 653 AA.

AC ABG23877:

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #23868.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS88064.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT biologically for genetic disorders or other traits and to assess

PT biodiversity

PS Claim 20; SEQ ID No 54236; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 653 AA;

XX Query Match 77.1%; Score 2473.5; DB 22; Length 653;

Best Local Similarity 75.5%; Pred. No. 3,9e-212;

Matches 493; Conservative 37; Mismatches 58; Indels 65; Gaps 2;

OY 1 MEARVYHALQKROVSLDVLGVNAGAPFLRYFAEETERCTFLANLADLGLVEELS 60  
 DB 1 MEARVYHALQKROVSLDVLGVNAGAPFLRYFAEETERCTFLANLADLGLVEELS 60  
 OY 61 ARGCRVSDPTGFLNPLTDDLLNKKLREELCGPFCVLPFQILLEPQIFRAE 120  
 DB 61 ARGCRVSDPTGFLNPLTDDLLNKKLREELCGPFCVLPFQILLEPQIFRAE 120  
 OY 121 LWRVDINDSPVFLDREITLNTLESTPGATFLDSAPDSVGINLNNTSSVYVPHI 180  
 DB 121 LWRVDINDSPVFLDREITLNTLESTPGATFLDSAPDSVGINLNNTSSVYVPHI 180  
 OY 181 NVHDGEGNVSELYLKVLDREVEPELRLTLTGDDGSPRSGTLLIRILVLDINDVP 240  
 DB 181 NVHDGEGNVSELYLKVLDREVEPELRLTLTGDDGSPRSGTLLIRILVLDINDVP 240  
 OY 241 EFVESLYKQVPENSPVGLVTVTSARDLDGSGNCEIYAFYATERLTKFRINSTSGN 300  
 DB 241 EFVESLYKQVPENSPVGLVTVTSARDLDGSGNCEIYAFYATERLTKFRINSTSGN 300  
 OY 301 LHLKALNVEALQVYTLTLTQAKDGGGLSGCTVVVHVVDINDNPPELLMSLTSPFENS 360  
 DB 301 LHLKALNVEALQVYTLTLTQAKDGGGLSGCTVVVHVVDINDNPPELLMSLTSPFENS 360  
 OY 361 PETVAVFRIRDRSGNNAKVCISQDLPLVLRPSVENVYTLTERALDREERTENIT 420  
 DB 361 PETVAVFRIRDRSGNNAKVCISQDLPLVLRPSVENVYTLTERALDREERTENIT 420  
 OY 421 ITVDTLGTPLRKTQNLVTVSDVNDNAPFESQTYTLRVENNSPLHIGSVATDRDS 480  
 DB 421 ITVDTLGTPLRKTQNLVTVSDVNDNAPFESQTYTLRVENNSPLHIGSVATDRDS 480  
 OY 481 GANA----- 484  
 DB 481 GANA----- 484  
 OY 485 GTMAOVISLPSQDPLPLASLVSINADNGHLFALSPWTTPMLRAFEFRNNPALHIGS 540  
 DB 485 GTMAOVISLPSQDPLPLASLVSINADNGHLFALSPWTTPMLRAFEFRNNPALHIGS 540  
 OY 541 VSTAETNIGVYSLPPRPPLPLASLVSINTDNGHLFALRSIDYELQEFERVGASDR 600  
 DB 541 VSTAETNIGVYSLPPRPPLPLASLVSINTDNGHLFALRSIDYELQEFERVGASDR 600  
 OY 537 GSPALSSQALVRLVADANDNAPFLVPLQNGSAPCTELVPR-AAEAGYLVAK 588  
 DB 601 GSTALSSSEALVRLVLDANDSSLFVLPPLQNGSAPCTELMPRAAERGGLVTK 653

## RESULT 5

ABG17157 ID ABG17157 standard; Protein; 797 AA.

AC ABG17157:

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #17148.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;  
 XX WPI: 2001-639362/73.  
 DR N-PSDB: AAS81344.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 20; SEQ ID No 47516; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 797 AA:

Query Match 71.8%; Score 2304; DB 22; Length 797;  
 Best Local Similarity 71.3%; Pred. No. 8e-197;  
 Matches 447; Conservative 77; Mismatches 103; Indels 0; Gaps 0;

QY 1 MEARVHAHQKROVSLCYFLGVSWAGAEPLRYFAETERTFLANLAIDGLGVEELS 60  
 DB 1 MEIRGALDRKROVLIFFVLGLSAGTESAHYVAEETIGSFVAMARLDGLGVEELS 60  
 QY 61 ARGCRIVSDETIGFLINPLTGLDLNLEKIDREELCGPTPEVLPOLLLEKFOIFPAE 120  
 DB 61 SREARVSDNDKKYHLDLTLNLEKIDREELCGPTPEVLPOLLLEKFOIFPAE 120  
 QY 121 LWRVDINDHSPVFLDREITLNLLESTTPGAFLLSEAHSDSGVGNLNRYTSSNVYFHI 180  
 DB 121 LCVKIDINDHSPFLDKETIKISEGTVGATFLMESADLDVGSNSLONYTSPSHFYI 180  
 QY 181 NVHNDGEGVYSEIVLADKLDRVEPELRALTTLGLDGSPPRSCTTLIRLVLINDVNP 240  
 DB 181 KIPDSSDKRIYELVLDRLDYEQEKRLTLTAADGSPKSGTTLVIRKVLINDNAP 240  
 QY 241 EEVESLYKYQVENSPPVGLVTVYVSGARDLDSNGEIIYAFYAFERTKTRPINSTGN 300  
 DB 241 EFPQSLYEQVPEDRPLSGWITISAKDLDAKNYKISTTFPHASEDIRKTEINPISGE 300  
 QY 301 LHLKAELEAIIQYTLTIQAKDGGGLSGCTVVVAVHTDINDNPELMLSSLTSPIPENS 360  
 DB 301 VNLRSPLDEVIQSYTINIQATDGGGLSGCTLLVYMDINDNPEVMTSITKRIPERA 360  
 QY 361 PEVYAVFETIRDRDSGNNAKWCSTODHLPEYLKPSVENEFTLVTERALDREKREYNT 420  
 DB 361 SETLVAFSLIDDSGDNRMICSTIDNLPFLKPFKNFTLVSEKALDRSQAQAYNT 420  
 QY 421 IVYTDGEPRLKTOHNLATVAVDVNDNAPTFSGOTYTLTVRRNNSPALHIGSVATDRDS 480  
 DB 421 IVYTDGEPRLKTEYNTVILSDVNDNAPTFGTSTITLTVRRNNSPALHIGSVATDRDS 480

QY 481 GANAQVYTSLLPPHPDPLVGSVLSINADNGOLFALRSIDFEALQAFERRVGADRGSPA 540  
 DB 481 GTNAQVYTSLLPPDRLPLASLVISINADNGHFLALRSIDYEALQEFERRVATDRGSPA 540  
 QY 541 LSSQALVRLVAVANDNAPVLYPLONGSAPCELPRAAEGLYAKVAVADGDSGQNA 600  
 DB 541 LSSBALVRLVLDANDLPVLYPLONGSAPCTELPRAAEPGLVTKVAVADGDSGQNA 600  
 QY 601 WLSYQLLKATEPGLFGVWANGGEVRTA 627  
 DB 601 WLSYQLLKATEPGLFGVWANGGEVRTA 627

RESULT 6  
 AAR86865  
 ID AAR86865 standard; Protein; 787 AA.

AC AAR86865;

XX 27-AUG-1996 (first entry)

DE Human protocadherin pc3.

KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;

KW catenin; therapy.

OS Homo sapiens.

PN W09600289-A1.

PD 04-JAN-1996.

XX 26-JUN-1995; 95WO-0508071.

XX 27-JUN-1994; 94US-0268161.

XX (DOHE-) DOHENY EYE INST.

PA Suzuki S;

XX WPI: 1996-068873/07.

DR N-PSDB: AAT03572.

PS Claim 15; Page 115-119; 146pp; English.

CC AAR86865-R86867 represent the sequences for three protocadherins. This

CC sequence represents the human protocadherin pc3. These sequences are

CC related to cadherin, and possess cell adhesive ability. Cadherins are

CC glycosylated integral membrane proteins that are involved in cell-cell

CC adhesion. Cadherins are composed of an N-terminal extracellular domain

CC which consists of 5 unique subdomains, a membrane spanning domain, and a

CC C-terminal cytoplasmic domain. The cytoplasmic domain interacts with

CC the cytoskeleton through catenins and other cytoskeleton associated

CC proteins. The cytoplasmic domain is not present in all cadherins, but in

CC those which possess it, it is essential for the cadherins adhesive

CC function. The cadherins which do not possess a cytoplasmic domain appear

CC to function via a different method from those with a cytoplasmic domain.

CC These sequences were isolated using primers 1 and 2 (see AAT03575 and

CC AAT03576) The proteins may have regulatory functions in the cell, as

CC well as the cell-cell adhesive properties. Antibodies produced against

CC these sequences are useful for modulating the binding a

CC protocadherins, and can be used therapeutically.

XX Sequence 787 AA;

QY Query Match 71.3%; Score 2287; DB 17; Length 787;  
 Best Local Similarity 72.3%; Pred. No. 2.6e-195;  
 Matches 447; Conservative 67; Mismatches 104; Indels 0

QY 10 OKROVSLCYFLGVSWAGAEPLRYFAETERTFLANLAIDGLGVEELS 60

FT	Modified-site	82..88	/note="N-myristoylation site"
FT	Modified-site	162..168	/note="N-myristoylation site"
FT	Modified-site	169..173	/note="N-myristoylation site"
FT	Modified-site	181..185	/note="N-glycosylation site"
FT	Modified-site	184..190	/note="N-glycosylation site"
FT	Modified-site	217..223	/note="N-myristoylation site"
FT	Modified-site	324..330	/note="N-myristoylation site"
FT	Modified-site	325..331	/note="N-myristoylation site"
FT	Modified-site	394..402	/note="N-myristoylation site"
FT	Modified-site	418..422	/note="Tyrosine kinase phosphorylation site"
FT	Modified-site	436..440	/note="N-glycosylation site"
FT	Modified-site	471..477	/note="N-glycosylation site"
FT	Modified-site	567..571	/note="N-myristoylation site"
FT	Modified-site	568..574	/note="N-glycosylation site"
FT	Modified-site	578..585	/note="N-myristoylation site"
FT	Modified-site	687..711	/note="Tyrosine kinase phosphorylation site"
FT	Modified-site	759..765	/note="transmembrane domain"
FT	Modified-site	781..785	/note="N-myristoylation site"
FT	Modified-site	788..792	/note="amidation site"
FT	Modified-site		/note="N-glycosylation site"
XX	WO200056889-A2.		
XX	28-SEP-2000.		
XX	01-MAR-2000; 2000WO-US05601.		
XX	23-MAR-1999; 99US-0125774.		
PR	23-MAR-1999; 99US-0125776.		
PR	24-MAR-1999; 99US-0125826.		
PR	31-MAR-1999; 99US-0127035.		
PR	05-APR-1999; 99US-0127706.		
PR	21-APR-1999; 99US-0130359.		
PR	27-APR-1999; 99US-0131270.		
PR	27-APR-1999; 99US-0131272.		
PR	27-APR-1999; 99US-0131291.		
PR	04-MAY-1999; 99US-0132371.		
PR	04-MAY-1999; 99US-0132379.		
PR	04-MAY-1999; 99US-0132383.		
PR	25-MAY-1999; 99US-0135750.		
PR	08-JUN-1999; 99US-0138166.		
PR	20-JUL-1999; 99US-0144791.		
PR	03-AUG-1999; 99US-0146970.		
PR	09-DEC-1999; 99US-0170262.		
XX	(GETH ) GENENTECH INC.		
XX	Desnoyers L, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J, Stewart TA, Watanabe CK, Wood WL, Zhang Z,		
XX	WPI; 2000-628263/60.		
XX	N-PDB; AAA96347.		

PR polypeptide and for therapeutic use  
XX  
PS Claim 12; Fig 24; 222pp; English.

CC The present sequence represents a secreted or transmembrane polypeptide.  
CC The specification describes polypeptides designated PRO1484, PRO4334,  
CC PRO1122, PRO1889, PRO1890, PRO1887, PRO1853, PRO4405,  
CC PRO4356, PRO4352, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,  
CC PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is  
CC useful for diagnosing tumour in a mammal. The polypeptides, their  
CC agonists and antagonists are useful treating a condition associated with  
CC expression or activity of the polypeptide. Conditions treated include  
CC obesity, diabetes or hyper- or hypo-insulinemia. The polypeptides are  
CC capable of inducing proliferation of mammalian kidney mesangial cells  
CC and are therefore useful for treating kidney disorders associated with  
CC decreased mesangial cell function such as Berger's disease or other  
CC nephropathies associated with Schönlein-Henoch purpura, celiac disease,  
CC dermatitis herpetiformis or Crohn's disease. The nucleic acids may be used  
CC to generate transgenic animals for use in development and screening of  
CC therapeutically useful reagents and also for chromosome identification  
CC and tissue typing.

XX Sequence 800 AA;

Query Match 70.1%; Score 2249; DB 21; Length 800;  
Best Local Similarity 70.4%; Pred. No. 6.7e-192;  
Matches 435; Conservative 70; Mismatches 113; Indels 0; Gaps 0;

QY 10 QKQVSLCVFLGVSWAGAPLRYFAETEGTFLANLADLGLCVETLSRGCRYSYD 69  
DB 10 RQKQVLEFLFVGWVSLAGSGFGRYSVTEETKGSFVYVNAKDLGALVAGTARGVSD 69  
QY 70 ETIGFLMLPDLGDLNKLKREELCGPEPCVLPOLLKPPQIYFAELWKEIINH 129  
DB 70 DNKQVLLSHLGNLTNKLREKLCGRKPEMLYFQILMDPFCQIYAEIRVNDINH 129  
QY 130 SPVFLDREITLNLSTPGATFLLESANDSVGINNLNNTISSNVYFHINWHDGECN 189  
DB 130 APFQDKKEYLYKISENTAGTAFRLERADDPGLGNGIGNYITSPSEFHINISGDEEM 189  
QY 190 VSELYLVKVDREVEPELRLITGLDGSPPRSCTTLIRILYLDINENVPFESLYKY 249  
DB 190 IYELVLDKALDREDEGELSLTLTALDGSPPRSCTTVRIYVLDVNDNAPQAGALYET 249  
QY 250 QVENSPPGSLVYVSARLDITSGNGEIVYAFYATERTLKPRIKSTSGNLKRELNY 309  
DB 250 QAPENSPGILYKVAEDVDSVNAEVSISFPDASENRTTFOIMPFGELFLRELIX 309  
QY 310 EAIQYTLTIOAKDGGGLSGCTVYVVDINDNPELMSLTSPDENSPETVAVFR 369  
DB 310 ELVNSKINIQAMDGGGLSARCVLEVLDINDNPELIVSSFSNVAENSPETPLAVFR 369  
QY 370 IRRDGSNNNAKWCSTODHLPFLKPSVENFTLVTERALDREKREYVITTYDGLTP 429  
DB 370 INRDSENGKWCYIOENLPFLKPSVENFTLVTERALDREKREYVITTYDGLTP 429  
QY 430 RLKQHLNLYTVSVNDNAPTFQCTYTLRVRENSPALHIGSVSADTDSGANAQVYS 489  
DB 430 RLKTEHNTIYLVSDVNDNAPAFQCTSYTLFVRKNSPALHIGSVSADTDSGANAQVYS 489  
QY 490 LLEPHDPLGLSLVSNADNGQLFALSLDFEALQAFERRVAGADRGSPALSSQALVRY 549  
DB 490 LLEPHDPLGLSLVSNADNGQLFALSLDFEALQAFERRVAGADRGSPALSSQALVRY 549  
QY 550 LVADANDNAFVLYPIQNGAPCTELVPRRAEAGYLVAKVYAVDGSQGANMSTYOLLA 609  
DB 550 LVADANDNSPVLVPIQNGAPCTELVPRRAEAGYLVAKVYAVDGSQGANMSTYOLLA 609  
QY 610 TERGLFGVMAHNGEVRTA 627  
DB 610 TERGLFGVMAHNGEVRTA 627

RESULT 8  
ID AAU29262 standard; Protein; 800 AA.  
XX AAU29262;  
AC 18-DEC-2001 (first entry)  
XX DE Human PRO polypeptide sequence #239.  
XX KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;  
KW dog; cat; pig; goat; rabbit; tumour; necrosis factor alpha; TNF-alpha;  
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.  
XX OS Homo sapiens.  
XX PN W0200168848-A2.  
XX PD 20-SEP-2001.  
XX PE 28-FEB-2001; 2001WO-US06520.  
XX PR 01-MAR-2000; 2000WO-US05601.  
PR 02-MAR-2000; 2000WO-US05841.  
PR 03-MAR-2000; 2000US-187202P.  
PR 06-MAR-2000; 2000US-186968P.  
PR 14-MAR-2000; 2000US-189320P.  
PR 14-MAR-2000; 2000US-189328P.  
PR 15-MAR-2000; 2000WO-US06884.  
PR 21-MAR-2000; 2000US-190828P.  
PR 21-MAR-2000; 2000US-191007P.  
PR 21-MAR-2000; 2000US-191048P.  
PR 21-MAR-2000; 2000US-191314P.  
PR 28-MAR-2000; 2000US-192655P.  
PR 29-MAR-2000; 2000US-193032P.  
PR 29-MAR-2000; 2000US-193033P.  
PR 30-MAR-2000; 2000WO-US08439.  
PR 04-APR-2000; 2000US-194449P.  
PR 04-APR-2000; 2000US-194647P.  
PR 11-APR-2000; 2000US-195975P.  
PR 11-APR-2000; 2000US-196000P.  
PR 11-APR-2000; 2000US-196187P.  
PR 11-APR-2000; 2000US-196690P.  
PR 11-APR-2000; 2000US-196820P.  
PR 18-APR-2000; 2000US-198121P.  
PR 18-APR-2000; 2000US-198585P.  
PR 25-APR-2000; 2000US-199397P.  
PR 25-APR-2000; 2000US-199550P.  
PR 25-APR-2000; 2000US-199654P.  
PR 03-MAY-2000; 2000US-201516P.  
PR 17-MAY-2000; 2000WO-US13705.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 30-MAY-2000; 2000WO-US14941.  
PR 02-JUN-2000; 2000WO-US15264.  
PR 05-JUN-2000; 2000US-209832P.  
PR 28-JUL-2000; 2000WO-US20710.  
PR 22-AUG-2000; 2000US-0644848.  
PR 24-AUG-2000; 2000WO-US23328.  
PR 08-NOV-2000; 2000WO-US30952.  
PR 01-DEC-2000; 2000WO-US32678.  
PR 20-DEC-2000; 2000WO-US34956.  
XX (GETH ) GENENTECH INC.  
XX PA Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
PI Pan J, Smith V, Watanabe CK, Wood WL, Zhang Z;  
XX WPI; 2001-602746/68.  
XX N-PSDB; AAS46163.  
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
PT presence of tumours, such as prostate and breast tumours, in mammals and

PT to screen for modulators of the compounds -

PS Claim 11; Fig 478; 774pp; English.  
 XY

CC Sequences AAU23024-AU239328 represent PRO polypeptides of the invention.  
CC The PRO polypeptides and their associated nucleic acids can be used to  
CC detect the presence of a tumour in a mammal by comparing the level of  
CC expression of a PRO polypeptide in a test sample of cells from the animal  
CC and a control sample of normal cells, whereby a higher level of  
CC expression in the test sample indicates the presence of a tumour in the  
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats  
CC and rabbits but are preferably human. The polypeptides can be used to  
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,  
CC when contacted with it. A specific polypeptide can be used to stimulate  
CC the proliferation or differentiation of chondrocyte cells. The PRO  
CC proteins can be used to determine the presence of tumours and also  
CC susceptibility to tumour development, particularly adrenal, lung, colon,  
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
CC can be used for genetic analysis of individuals with genetic disorders.  
CX

**SQ Sequence 800 AA;**

Query Match	70.1%;	Score 2249;	DB 22;	Length 800;
Best Local Similarity	70.4%;	Pred. No. 6.7e-192;		
Matches 435;	Conservative 70;	Mismatches 113;	Indels 0;	Gaps 0

[illegible]

## RESULT 9

ABG23214  
ID ABG23214 standard; Protein; 1888 AA  
vv

AC ABG23214; .  
xy

DT 18-FEB-2002 (first entry)  
yy

Novel human diagnostic protein #233205

KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food quality control

[illegible]

XX  
DN

W0300175067-2

XX  
DN 11-OCT-2001

XX  
PE 30-MAR-2001. 2001WO-TTS08631

XX  
PR 31-MAR-2000. 2000HS-0540217

PR 23-AUG-2000; 200005-0649167.  
XX

PA (HYSE-) HYSEQ INC.  
XX

P1 Drmanac KJ, Liu C, Tang YT, XX

DR WPI; 2001-639362/13.  
DR N-PSDB: AAS87401.

AA	PT	New isolated polynucleotide
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

PT responsible for a

Universally -

XX  
Citation 20; SEQ ID NO 553/3; 103bp; ENGLISH:

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy, technology to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG03037 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences).

SQ Sequence 1888 AA;

Query Match	Score	DB	Length
Best Local Similarity	71.18;	Pred. No. 3.3e-190;	
Matches 441; Conservative	69;	Mismatches 108;	Indels 2; Gaps 2

10 QKROSLICVLEGSMAQAPLRYPAAEERTPTLANAIDIGISVEELASRCRIVSD 69  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
26 EQOVLILLILLEETLAGWEPRRYSMEETEKSFANLANDGIGVELAKERARAYSE 85  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

70 ETIGLLINPLTGLLINKEKLDEEELCGPPEBCVLPOLLLEKPQIFRAELAWINDH 129  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

86 DNEGGALDLDGTGLLINEKLDEKLCGPTEPCIMHFVLLKKPLEYFRALVTLDINDH 145  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|





BR 11-SEP-1998; 98US-0100038.  
 XX (GETH ) GENENTECH INC.  
 XX  
 XX Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;  
 XX WPI: 1999-551358/46.  
 DR N-PSDB; AAZ34233.  
 XX  
 XX New secreted and transmembrane polypeptides and their polynucleotides,  
 PT useful for treating blood coagulation disorders, cancers and cellular  
 PT adhesion disorders -  
 XX  
 XX Claim 12; Fig 163; 530pp; English.  
 XX  
 XX The present invention describes secreted and transmembrane polypeptides  
 CC and their polynucleotides. The nucleotide sequences are useful as  
 CC sources of probes, primers, for chromosome mapping, and for generation  
 CC of antisense sequences. They can also be used to create transgenic  
 CC animals. The proteins can be used to treat a variety of diseases and  
 CC disorders, depending on their function. Diseases that may be treated  
 CC include blood coagulation disorders, cancers and cellular adhesion  
 CC disorders. They may also be used to raise antibodies. AAZ33891 to  
 CC AAZ3438, and AA41685 to AA41774 represent polynucleotide and  
 CC polypeptide sequence given in the exemplification of the present  
 CC invention.  
 XX  
 XX Sequence 798 AA;  
 SQ  
 Query Match 66.8%; Score 2142; DB 20; Length 798;  
 Best Local Similarity 68.0%; Pred. No. 2.5e-182;  
 Matches 427; Conservative 66; Mismatches 133; Indels 2; Gaps 2;  
 1 MEANVHALKQVSLLCVFLGVSMAW-ABPLRYFAEETRGFTLANLAIDGLVEEL 59  
 1 MEASGKLICRQVLFSSLLGLSLAGAEPRYSVEETGSEFYNLAKDGLRQER 60  
 60 SARCRIVDEITGFLNPLTGLLNEKLDRELCPPEPCVPLQLLEKPFOTFRA 119  
 61 SRGRVAVSRKCNKHLQNLQETADLLNEKLDRELDCGHTPEVLRQVLLSEPFEEFQA 120  
 120 ELAVRDNDHSFVLDREITLNIESTTTPGATFELLESABSDGGINLRYTSSNVYFH 179  
 121 ELQYIDINDHSFVLDKQMLVKVSESSPPTTPPKAELDVGQNNIENYITSPSYFR 180  
 180 INVHNGEGVNYSELVLDKVDREVEPELRLTGLDGGSPRSGTTLRIIVLDINDY 239  
 181 VLRKRSDGRRKPELVLDKALDREERELRLTLTALDGGSPRSGTAQVYIEVLDVND 240  
 240 PEPFESLYKQVSPVSVLVYTSARDLDGNSGNIYAFYATERTKTRINSTG 299  
 241 PEFQPTKRVQISDSVGLVVKVSTVDVGVNGEISYSLQASSEETGKPKINPLTG 300  
 300 NLHLKELANVETIQTTLTQAKDGGISGCTVVVHTDINNPELMSLTSPFIPEN 359  
 301 EIELKRLDEKIQSYVNIENARDAGTFSKCTVLIQVIDVNDHAEVMSATTSIFEN 360  
 360 SPETVAVRIRDRDSGNNAKWCISODHLPEVLKPSVENEYTLVTERALDREERTYNI 419  
 361 APETVAVLFVSVDLDSGNGKISQIEDLPFLK-SAEKFYTLTERPLDRSREAYNI 419  
 420 TIIVTDGTRPKTQHLVTVTSVDVNDNAFTESQTYTLKRVNNNSPALHIGSATDND 479  
 420 TIIVTDGTRPKTQHLVTVTSVDVNDNAFTESQTYTLKRVNNNSPALHIGSATDND 479  
 480 SGANAQVTSILRPPHPOLPLGLSVGINADNGQLFALRSIDFALDAFEFRVGAARSGP 539  
 480 SGANAQVTSILRPPHPOLPLGLSVGINADNGQLFALRSIDFALDAFEFRVGAARSGP 539  
 540 ALSQALVRLVADANDNAPFVLYPLQNSAPCTELVPRAAEGYLVARVAVDGGSCN 599  
 540 ALSSEALVRLVADANDNSPFVLYPLQNSAPCTELVPRAAEGYLVARVAVDGGSCN 599

QY 600 AMLSYQLKATEPGLFGVAHNGEVRTA 627  
 DB 600 AMLSYQLKATEPGLFGVAHNGEVRTA 627  
 RESULT 11  
 AAB44302  
 ID AAB44302 standard; Protein; 798 AA.  
 AC AAB44302;  
 XX  
 XX 08-FEB-2001 (first entry)  
 DT  
 XX Human PRO531 (UNQ332) protein sequence SEQ ID NO:405.  
 DE  
 XX Human; secreted protein; transmembrane protein; PRO; EST; cytosolic;  
 KW expressed sequence tag; detection; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200053756-A2.  
 PN  
 XX 14-SEP-2000.  
 PD  
 XX 18-FEB-2000; 2000WO-US04341.  
 PF  
 XX 08-MAR-1999; 99WO-US05028.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 29-MAR-1999; 99US-0126773.  
 PR 21-APR-1999; 99US-0130232.  
 PR 28-APR-1999; 99US-0131445.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99US-0528313.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28565.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 30-DEC-1999; 99WO-US31243.  
 PR 30-DEC-1999; 99WO-US31274.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 XX  
 XX (GETH ) GENENTECH INC.  
 XX  
 XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
 PI Ferrata N, Filvaroff E, Fong S, Gerber H, Gerritsen ME;  
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
 PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;  
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;  
 DR WPI: 2000-611443/58.  
 DR N-PSDB; AAC78558.  
 XX  
 XX Novel PRO polypeptides and polynucleotides used in detection methods,  
 PT to target bioactive molecules to specific cells, and to modulate  
 PT cellular activities -  
 XX  
 XX Claim 12; Fig 163; 636pp; English.  
 PS  
 XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed  
 CC sequence tag) sequences which encode secreted or transmembrane PRO  
 CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic  
 CC activity. The polynucleotides and polypeptides can be used for detecting  
 CC the presence of PRO polypeptides in samples, for linking bioactive  
 CC molecules to cells and for modulating biological activities of cells,  
 CC using the polypeptides for specific targeting. The polypeptide targeting  
 CC can be used to kill the target cells, e.g. for the treatment of cancers.  
 CC The polypeptide pairs provide specific targeting of bioactive molecules  
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in  
 CC the isolation of the PRO polynucleotide sequences.





Db 301 EIELKQIDPEKLOSIEYENIEARDAGTFSGKCTVLIQVIDVNDHAPVNTSAFTSPLEN 360  
 QY 360 SPETVAVAFRIIDRDGNNAMKCSIODHLPFLVKSPVENFYTLVTRALDREERTFYNI 419  
 Db 361 APEITVALFVSDDSDGSKISCSIEDLPFLK -SAENFYTLITRPLDRESRAFYNI 419  
 QY 420 TITVTDICTPRKQNHUITYVSVYDNNAPFESQTYTTLVRENNSPALHIGSVATDRD 479  
 Db 420 TITVTDICTPRKQNHUITYVSVYDNNAPFESQTYTTLVRENNSPALHIGSVATDRD 479  
 QY 480 SGANAQVTSLSLPHDPOLPLGSLVSNADNGOLFALRSIDFEALAEFEERVGAAADRGSP 539  
 Db 480 SGTMAQVTSLSLPHDPOLPLGSLVSNADNGOLFALRSIDFEALAEFEERVGAAADRGSP 539  
 QY 540 ALSSQALVRLVADANDNAFVLYPLQNGSAPCTELVPRAAGVYLVAKVAVADGSGON 599  
 Db 540 ALSSQALVRLVADANDNAFVLYPLQNGSAPCTELVPRAAGVYLVAKVAVADGSGON 599  
 QY 600 AMLSYQLKATELGFEGVMAHNEVFTA 627  
 Db 600 AMLSYQLKATELGFEGVMAHNEVFTA 627

## RESULT 13

AAB33433

ID AAB33433 standard; Protein; 798 AA.

AC AAB33433:

DT 29-JAN-2001 (first entry)

DE Human PRO531 protein UNQ332 SEQ ID NO:107.

XX Human; immune related disease; diagnosis; antinflammatory; cardiant;  
 KW dermatological; antiarthritic; antirheumatic; immunosuppressive;  
 KW haemostatic; antithyroid; antidiabetic; neurotropic; neuroprotective;  
 KW antianemic; hepatotropic; virucide; antipsoriatic; antiallergic;  
 KW antiasthmatic; systemic lupus erythematosus; rheumatoid arthritis;  
 KW osteoarthritis; spondyloarthritis; systemic sclerosis; sarcoidosis;  
 KW idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;  
 KW systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;  
 KW autoimmune thrombocytopenia; immune-mediated renal disease;  
 KW demyelinating disease; hepatobiliary disease; Whipple's disease;  
 KW inflammatory bowel disease; gluten-sensitive enteropathy;  
 KW autoimmune disease; immune-mediated skin disease; allergic disease;  
 KW immunological disease; transplantation associated disease;  
 KW graft rejection; graft-versus-host-disease.

XX Homo sapiens.

XX WO200053758-A2.

PD 14-SEP-2000.

PF 02-MAR-2000; 2000WO-US05841.

PR 08-MAR-1999; 99WO-US05028.  
 PR 10-MAR-1999; 99US-0123618.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 23-MAR-1999; 99US-0123775.  
 PR 12-APR-1999; 99US-0128849.  
 PR 20-APR-1999; 99WO-US08615.  
 PR 28-APR-1999; 99US-0131445.  
 PR 04-MAY-1999; 99US-0133371.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 28-JUL-1999; 99US-0146222.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 13-SEP-1999; 99WO-US20944.

PR 15-SEP-1999; 99WO-US21090.  
 PR 15-SEP-1999; 99WO-US21547.  
 PR 05-OCT-1999; 99WO-US23089.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 29-NOV-1999; 99WO-US28214.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 01-DEC-1999; 99WO-US28409.  
 PR 01-DEC-1999; 99WO-US28301.  
 PR 02-DEC-1999; 99WO-US28634.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28564.  
 PR 02-DEC-1999; 99WO-US28565.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 20-DEC-1999; 99WO-US30999.  
 PR 30-DEC-1999; 99WO-US31274.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 PR 11-FEB-2000; 2000WO-US03565.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 22-FEB-2000; 2000WO-US04414.

(GETH ) GENENTECH INC.

PA Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W,  
 PI Kabakoff RC, Lu Y, Pan J, Penica D, Shelton DL, Smith V,  
 PI Stewart TA, Tumas D, Watanabe CK, Wood WI, Yan M,

WPI; 2000-572271/53.

DR N-PSDB; AAC58598.

PT Sixty four PRO polypeptides, useful in the diagnosis and treatment of  
 PR immune related disorders, e.g. systemic lupus erythematosus, rheumatoid  
 PR arthritis, osteoarthritis, thyroiditis and diabetes mellitus -  
 XX Claim 33; Fig 40; 309pp; English.

XX The present invention describes sixty four human PRO proteins which can  
 CC be used in the treatment of immune related diseases. The human PRO  
 CC proteins, anti-PRO antibodies, agonists and antagonists are useful for  
 CC treating and diagnosing immune related disorders. The disorders are  
 CC selected from systemic lupus erythematosus, rheumatoid arthritis,  
 CC osteoarthritis, juvenile chronic arthritis, spondyloarthritis,  
 CC systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's  
 CC syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic  
 CC anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus,  
 CC immune-mediated renal disease, demyelinating diseases of the central  
 CC and peripheral nervous systems, hepatobiliary diseases, inflammatory  
 CC bowel disease, gluten-sensitive enteropathy and Whipple's disease,  
 CC autoimmune or immune-mediated skin diseases, allergic diseases,  
 CC immunological diseases of the lung, and transplantation associated  
 CC diseases including graft rejection and graft-versus-host-disease.  
 CC AAC58397 to AAC58578 represent PCR primers and hybridisation probes used  
 CC in the isolation of human PRO sequences. AAC58579 to AAC58642 and  
 CC AAB33414 to AAB33477 represent human PRO polynucleotide and protein  
 CC sequences given in the exemplification of the present invention.

SQ Sequence 798 AA;

Query Match 66.8%; Score 2142; DB 21; Length 798;  
 Best Local Similarity 68.0%; Pred. No. 2,5e-182;  
 Matches 427; Conservative 66; Mismatches 133; Indels 2; Gaps 2;

QY 1 MEARVVAHAKROVSLICVFGVSMAG-ABPLRYFAAEETERTPLANTLADLGVEEL 59  
 Db 1 MEASGKILCRQRYLPSFLLGLSLGACAPRPRSYVEETEGSSFTYNLAKDLGQREF 60  
 QY 60 SARCGRIVSDETIGFLLNPLVGDLLNEXLDEELCGPTPCVLPFOLLKPKQIFRA 119  
 Db 61 SRGQVRYVSRGNKLHLQINQETADLLNEXLDEELCGHTPCVLPFOLLKPKQIFRA 120  
 QY 120 ELWVRDINDHSVFLDREITLNLLESTTPGATFLLESANDSDVGINNLNNTISSNVYH 179

```

Db      121 ELQVINDHDSVFLDKOMLVKVSSESPGCTFFPLKNAEDDVCGNNIENYIISPNSYFR 180
Qy      180 INVHNGSGNYSSELYDKVLDREVPRLRTLTGLDGGSPRRSTTLIRLVLDINDV 239
Db      181 VLTFRRSRGRKPELVLDKALDREBEALRTLTALDGGSPRRSTTAQVYEVLDVNDNA 240
Qy      240 PEFVSLKVOVPENSPPGSLVTVVSARDLDTGSGEIVAFVFAETERTLTFRINSTG 299
Db      241 PEFQPFYRVQISDESPGFLVVKVKSATDVDTGVNGEISTSLFQASSEIGTFTKINPLTG 300
Qy      300 NLHLKAEINYEAIQTYTLTIOAKDGGISGKCTVYVHVHTDINDNPPELLMSLTSPIDEN 359
Db      301 ETELKQDLDFEKLQSYEVNIEARDAFTSGKCTVLIQVIVNDNAPVETMSAFTSPIDEN 360
Qy      360 SPETVVAVFRIRDRDSGNNAKAVCSIODHLPVLPKPSVENYTLTVERALDREERTENI 419
Db      361 ABEYVALEFSVSDLDGSENGKISCSIOEDLPFLK-SAENYTLTTERPLDRESAEYNI 419
Qy      420 TTVVDLGTPLRKTQHNLTVTVDVNDNAPFESQTYTLTVERENNSPALHIGVSATDRD 479
Db      420 TTVVDLGTPLRKTQHNLTVTVDVNDNAPFESQTYTLTVERENNSPALHIGVSATDRD 479
Qy      480 SCANQVTVSLIPHPDPLGSLVSINADNGQLFALNSLDEALQAFEFVGAADRGSP 539
Db      480 SCTNAGVTVSLIPHPDPLGSLVSINADNGQLFALNSLDEALQAFEFVGAADRGSP 539
Qy      540 ALSALVAVIVADANDNAPFLVPLONGSAPCTELVRAEAGVIVAKVYVAVDGDGSON 599
Db      540 ALSSELVAVIVADANDNAPFLVPLONGSAPCTELVRAEAGVIVAKVYVAVDGDGSON 599
Qy      600 AMLSYQLKATEPGLGEGVVAHNGEVRTA 627
Db      600 AMLSYQLKATEPGLGEGVVAHNGEVRTA 627

```

## RESULT 14

AAU83657

ID AAU83657 standard; Protein: 798 AA.

XX AAU83657;

XX 08-MAY-2002 (first entry)

XX Human PRO protein, Seq ID No 132.

XX Human; secreted protein; PRO; tumour; lung cancer; colon cancer;

XX breast cancer; prostate tumour; rectal tumour; liver tumour;

XX pericyte cell proliferation; chondrocyte cell proliferation;

XX tumour necrosis factor-alpha.

XX Homo sapiens.

XX WO200208286-A2.

XX 31-JAN-2002.

XX 29-JUN-2001; 2001WO-US21066.

XX 20-JUL-2000; 2000US-219556P.

XX 25-JUL-2000; 2000US-220585P.

XX 25-JUL-2000; 2000US-220605P.

XX 25-JUL-2000; 2000US-220607P.

XX 25-JUL-2000; 2000US-220624P.

XX 25-JUL-2000; 2000US-220638P.

XX 25-JUL-2000; 2000US-220664P.

XX 25-JUL-2000; 2000US-220666P.

XX 26-JUL-2000; 2000US-220893P.

XX 28-JUL-2000; 2000WO-US20710.

XX 23-AUG-2000; 2000WO-US23522.

XX 24-AUG-2000; 2000WO-US23328.

XX 15-SEP-2000; 2000US-000000P.

XX 10-NOV-2000; 2000WO-US30873.

PR

```

PR      28-NOV-2000; 2000US-253646P.
PR      01-DEC-2000; 2000WO-US32678.
PR      20-DEC-2000; 2000US-0747259.
PR      20-DEC-2000; 2000WO-US34956.
PR      28-FEB-2001; 2001WO-US06520.
PR      10-MAY-2001; 2001US-0854280.
PR      25-MAY-2001; 2001WO-US17092.
XX      (GETH ) GENENTECH INC.
PA      Baker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;
XX      Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WI;
XX      WPI; 2002-172001/22.
DR      N-PSDB; ABR33601.
XX      One hundred and twenty two nucleic acids encoding PRO polypeptides,
PT      useful for treating a PRO related disorder and for diagnosing tumours
PT      such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal
PT      tumour or liver tumour -
XX      Claim 11; Figure 132; 359pp; English.
XX      The invention relates to one hundred and twenty two nucleic acids
XX      encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides
XX      encode human secreted proteins. The PRO nucleic acids, polypeptides,
XX      agonists and antagonists are useful for treating a PRO related disorder.
XX      The PRO polypeptides are useful for diagnosing tumours, especially lung
XX      cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or
XX      liver tumour. The PRO polypeptides are useful for stimulating the
XX      proliferation of, or gene expression, in pericyte cells, for stimulating
XX      the proliferation or differentiation of chondrocyte cells, for
XX      stimulating the release of tumour necrosis factor-alpha from human dermal
XX      for stimulating or inhibiting the proliferation of normal human dermal
XX      fibroblast cells. The PRO polypeptide may also be used as molecular
XX      weight markers and for tissue typing. The PRO nucleic acids have
XX      applications in molecular biology, including use as hybridisation probes,
XX      and in chromosome and gene mapping. AAU83592-AAU83713 represent human PRO
XX      protein sequences of the invention.
SO      Sequence 798 AA;
Query Match 66.8%; Score 2142; DB 23; Length 798;
Best Local Similarity 68.0%; Pred. No. 2,5e-182;
Matches 427; Conservative 66; Mismatches 133; Indels 2; Gaps 2;
Qy      1 MEARVVAHLQKROYSLCLVFLGVSMAG-APRLRFVAEFERGFPLNLANIDIGLGEEL 59
Db      1 MEASGKLICRQROYLFSFLGLSLAGAAEPRSYVVEETEGSFVNLAKDGLGEORRF 60
Qy      60 SARGCRIVSDEITIGFLINPLTGDLLNKKLDRRELCPTEPCVLPQDLLEKPFQIFRA 119
Db      61 SRGCVRVYSRGNKHLQNLQNOETADLLNKKLDRREDLGGHTEPCVLRQVYLLESFEEFQA 120
Qy      120 ELWVRDINDHSPVFLDREITNILESTPGATFLLESAAHSDVGINNLRNNTYSSNVEFH 179
Db      121 ELQVINDHDSVFLDKOMLVKVSSESPGCTFFPLKNAEDDVCGNNIENYIISPNSYFR 180
Qy      180 INVHNGSGNYSSELYDKVLDREVPRLRTLTGLDGGSPRRSTTLIRLVLDINDV 239
Db      181 VLTFRRSRGRKPELVLDKALDREBEALRTLTALDGGSPRRSTTAQVYEVLDVNDNA 240
Qy      240 PEFVSLKVOVPENSPPGSLVTVVSARDLDTGSGEIVAFVFAETERTLTFRINSTG 299
Db      241 PEFQPFYRVQISDESPGFLVVKVKSATDVDTGVNGEISTSLFQASSEIGTFTKINPLTG 300
Qy      300 NLHLKAEINYEAIQTYTLTIOAKDGGISGKCTVYVHVHTDINDNPPELLMSLTSPIDEN 359
Db      301 ETELKQDLDFEKLQSYEVNIEARDAFTSGKCTVLIQVIVNDNAPVETMSAFTSPIDEN 360
Qy      360 SPETVVAVFRIRDRDSGNNAKAVCSIODHLPVLPKPSVENYTLTVERALDREERTENI 419
Db      361 ABEYVALEFSVSDLDGSENGKISCSIOEDLPFLK-SAENYTLTTERPLDRESAEYNI 419

```



70 FTIGFI.I.I.NPI.TGDI.I.NEKI.DREEI.GPTEPCVI.PFOILL.EKPEOIFBAELWVBDINDH 129

```

; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 787 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-268-161A-110

Query Match          71.3%  Score 2287; DB 1; Length 787;
Best Local Similarity 72.3%; Pred. No. 2.4e-207;
Matches 447; Conservative 67; Mismatches 104; Indels 0; Gaps 0;

130 SPVFLDREITNLESTPGATFLESADSDVGINNLRYNTISSNYEHIINVDNGRN 129
130 SPEPEREMTKIPETSSLGIVFLKARDDVGSNNVQNTNISPNSHFVSTRKGDGR 189
130 YSELVLDKVIDREVPBELRLTLGLDGSPPRSCTTLIRILVLDINDNPEVESLYKV 249
190 KPELVLDTELDREDAELRLTLAVDGSPPRGCTVOILLVLDANDNAPEVQALYEV 249
250 QVPENSPVGSIVYVSARDLDTGNGELVYAFVATEETLTKTRINSGNLHKAELNY 309
250 QVPENSPVGSIVYVSARDLDTGNGELVYAFVATEETLTKTRINSGNLHKAELNY 309
310 EIQTYYTLTIOAKDGGGLSGKCTVYVHTDINDNPELLMSLSLSPENSPETVAVFR 369
310 ETMSYDLDIEASDGGGLSGKCSYKVLVDNDNPELSLSSLPENSPETVAVFR 369
370 IRDSDGNNAKMCVSIQDHLPEVLPKPSVENFYTLTERALDREERTENITITVDLGP 429
370 IRDSDGNNAKMCVSIQDHLPEVLPKPSVENFYTLTERALDREERTENITITVDLGP 429
430 RLKTOHNLTVYSDVNDNAPFESOTTYTLRVRENNSPALHIGVSATDRDGSANAQVYS 489
430 RLKTOHNLTVYSDVNDNAPFESOTTYTLRVRENNSPALHIGVSATDRDGSANAQVYS 489
490 LPPHDPQLPLGSLVSIINADNGQLFALRSIDFEALQAFEFVGAAGRSFALSSQALVRV 549
490 LPPHDPQLPLGSLVSIINADNGQLFALRSIDFEALQAFEFVGAAGRSFALSSQALVRV 549
550 LVADANDNAPVLYPLONGSAPCTELVPRAEAGYLVAKVAVDGSQGNAMLSYOLKA 609
550 LVADANDNAPVLYPLONGSAPCTELVPRAEAGYLVAKVAVDGSQGNAMLSYOLKA 609
610 TERGLFGVMAHNGEVRTA 627
610 TERGLFGVMAHNGEVRTA 627

; RESULT 2
US-08-268-161A-110
; Sequence 110, Application US/08268161A
; Patent No. 5798224
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/268,161A
; FILING DATE: June 27, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,337
; REFERENCE/DOCKET NUMBER: 27866/32149
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856

```

```

; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 787 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-268-161A-110

Query Match          71.3%  Score 2287; DB 1; Length 787;
Best Local Similarity 72.3%; Pred. No. 2.4e-207;
Matches 447; Conservative 67; Mismatches 104; Indels 0; Gaps 0;

130 SPVFLDREITNLESTPGATFLESADSDVGINNLRYNTISSNYEHIINVDNGRN 189
130 SPEPEREMTKIPETSSLGIVFLKARDDVGSNNVQNTNISPNSHFVSTRKGDGR 189
190 YSELVLDKVIDREVPBELRLTLGLDGSPPRSCTTLIRILVLDINDNPEVESLYKV 249
190 YSELVLDKVIDREVPBELRLTLGLDGSPPRSCTTLIRILVLDINDNPEVESLYKV 249
250 QVPENSPVGSIVYVSARDLDTGNGELVYAFVATEETLTKTRINSGNLHKAELNY 309
250 QVPENSPVGSIVYVSARDLDTGNGELVYAFVATEETLTKTRINSGNLHKAELNY 309
310 EIQTYYTLTIOAKDGGGLSGKCTVYVHTDINDNPELLMSLSLSPENSPETVAVFR 369
310 ETMSYDLDIEASDGGGLSGKCSYKVLVDNDNPELSLSSLPENSPETVAVFR 369
370 IRDSDGNNAKMCVSIQDHLPEVLPKPSVENFYTLTERALDREERTENITITVDLGP 429
370 IRDSDGNNAKMCVSIQDHLPEVLPKPSVENFYTLTERALDREERTENITITVDLGP 429
430 RLKTOHNLTVYSDVNDNAPFESOTTYTLRVRENNSPALHIGVSATDRDGSANAQVYS 489
430 RLKTOHNLTVYSDVNDNAPFESOTTYTLRVRENNSPALHIGVSATDRDGSANAQVYS 489
490 LPPHDPQLPLGSLVSIINADNGQLFALRSIDFEALQAFEFVGAAGRSFALSSQALVRV 549
490 LPPHDPQLPLGSLVSIINADNGQLFALRSIDFEALQAFEFVGAAGRSFALSSQALVRV 549
550 LVADANDNAPVLYPLONGSAPCTELVPRAEAGYLVAKVAVDGSQGNAMLSYOLKA 609
550 LVADANDNAPVLYPLONGSAPCTELVPRAEAGYLVAKVAVDGSQGNAMLSYOLKA 609
610 TERGLFGVMAHNGEVRTA 627
610 TERGLFGVMAHNGEVRTA 627

; RESULT 3
US-08-453-702A-110
; Sequence 110, Application US/08453702A
; Patent No. 5891706
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606

```





Db 370 IRDRSGNGKMCISIODVFPKLPSENFYRLVTEGALDRETRAENITITITDGT 429  
 QY 430 RLKTOHNLTVSDVNDNAPFESQTTTLRYRENNSPALHIGVSATDRSGANAQVYS 489  
 Db 430 RLKTEQSTIVLSDVNDNAPAFOTSTYTLFVRENNSPALHIGVSATDRSGANAQVYS 489  
 QY 490 LLPHPDPLGSLVSINADNGQLFALRSLDFEALQAFEEFVGAADRGSPALSSQALVRY 549  
 Db 490 LLPHPDPLGSLVSINADNGQLFALRSLDFEALQAFEEFVGAADRGSPALSSQALVRY 549  
 QY 550 LVADANDAPFVLPYPLONGSAPCTELVPRAEAGVYAKVAVDGDGSONAMLSYOLLA 609  
 Db 550 LVADANDAPFVLPYPLONGSAPCTELVPRAEAGVYAKVAVDGDGSONAMLSYOLLA 609  
 QY 610 TEPGLFGVMAHNGEVRTA 627  
 Db 610 TEPGLFGVMAHNGEVRTA 627

## RESULT 5

PCT-US95-08071-110  
 ; Sequence 110, Application PC/TUS9508071  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 115  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
 ; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/08071  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/12588  
 ; FILING DATE: 23 DEC 1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/998,003  
 ; FILING DATE: 29 DEC 1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Noland, Greta E.  
 ; REGISTRATION NUMBER: 35,302  
 ; REFERENCE/DOCKET NUMBER: 32149  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 110:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 787 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; PCT-US95-08071-110

## Query Match

Best Local Similarity 72.3%; Pred. No. 2,4e-207;  
 Matches 447; Conservative 67; Mismatches 104; Indels 0; Gaps 0;

QY 10 QKROVSLLEVLGVSWAGAPLRYVAETERTGLANLAIDLGLGVLELSARGCRIYSD 69  
 Db 10 EOROVLLILLLEVLGAWEPFRYSYMEETRGESFVANLANLDGLGVLELSARGCARVSE 69

QY 70 ETIGLELLPLDGLLNEKLDREELCGTPECVLPFOLLEKPPQIRAEIWRVINDH 129  
 Db 70 DNEQGLDQLDQLGLINKLREKLCGTEPCIMHFPVLLKPLVEFRAELLYVDINDH 129  
 QY 130 SPVFLDREITLNLILSTPGATFELLESADSDVGINLNRYTSSNVFHIIVHNGEGN 189  
 Db 130 SPEEPRREMTIKIPETSLIGVFPPLKKAADLDVGSNNQVNNYISNHFHSTRRGGR 189  
 QY 190 VSELYLVKVDREVPELRLTLGLDGSPPRSGTTLRIIVYDINDNVEPEVSLYK 249  
 Db 190 KYPELVLDLELREQAEIRLTLTAVDGSSPPRSGTVOILLIVDANDNAEVEFVALYEV 249  
 QY 250 QVPENSPVSLVYVSARDLDTGSNGEIVYAFATEERTLTERINSTGNHLKAEIYN 309  
 Db 250 QVPENSPVSLVYKVSARDLDTGTNGEISLYYSQETIDKPFELSSLSGEIRLIKLD 309  
 QY 310 EAIQTYTLTQAKDGGSLGKCTVVVHTDINDNPELLMSLSPIPENSPEVAVFR 369  
 Db 310 ETMSYDIDIEASDGGSLGKCSVSVKYLDVNDNPELISLSITSPINSPETEVALLR 369  
 QY 370 IRDRSGNNAKVCSTODHLPVLPKPSYENFTYTLTERALDREKTEENITITDGT 429  
 Db 370 IRDRSGNGKMCISIODVFPKLPSENFYRLVTEGALDRETRAENITITITDGT 429  
 QY 430 RLKTOHNLTVSDVNDNAPFESQTTTLRYRENNSPALHIGVSATDRSGANAQVYS 489  
 Db 430 RLKTEQSTIVLSDVNDNAPAFOTSTYTLFVRENNSPALHIGVSATDRSGANAQVYS 489  
 QY 490 LLPHPDPLGSLVSINADNGQLFALRSLDFEALQAFEEFVGAADRGSPALSSQALVRY 549  
 Db 490 LLPHPDPLGSLVSINADNGQLFALRSLDFEALQAFEEFVGAADRGSPALSSQALVRY 549  
 QY 550 LVADANDAPFVLPYPLONGSAPCTELVPRAEAGVYAKVAVDGDGSONAMLSYOLLA 609  
 Db 550 LVADANDAPFVLPYPLONGSAPCTELVPRAEAGVYAKVAVDGDGSONAMLSYOLLA 609  
 QY 610 TEPGLFGVMAHNGEVRTA 627  
 Db 610 TEPGLFGVMAHNGEVRTA 627

## RESULT 6

US-08-453-695A-112  
 ; Sequence 112, Application US/08453695A  
 ; Patent No. 5708143  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 115  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
 ; STREET: 233 South Wacker, 6300 Sears Tower  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/453,695A  
 ; FILING DATE:  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: No. 5708143and, Greta E.  
 ; REGISTRATION NUMBER: 35,302  
 ; REFERENCE/DOCKET NUMBER: 32658  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300





Db 599 AWLSFOLKATEPGLFSVAHNGEVRT 625

RESULT 8  
US-08-453-702A-112  
Sequence 112 Application US/08453702A  
Patent No. 3891706  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSEE: Borun  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,702A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5891706and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32657  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 112:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 797 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-702A-112

Query Match 62.6%; Score 2009.5; DB 2; Length 797;  
Best Local Similarity 62.7%; Pred. No. 4.5e-181;  
Matches 393; Conservative 83; Mismatches 148; Indels 3; Gaps 3;

Db 1 MEARVVALORQVSLVFLVGSW-AGAEPRLRYFAETERGFLNLAIDGLGYEEL 59  
1 METALAKIPOOROVFLTLT-LSLWKSSSEAIRSMPEETESGYMANLAKDGLRIGEL 59

Db 60 SARCGRIYSDETIGFLNPLTGLDLNLEKIDREELCGPTEPCVLPOLLERKFOIFRA 119  
60 SSRGAQIHYKGNKELLQDAETGNLFLKELKIDREELCGTEPCVLPOLLERKFOIFRA 119

Db 120 ELWRDINDSPVLDREITLNLIEESTTPGATFLLESAHSDVGINLKNYTTSSNVEYH 179  
120 ELQITDINDSPVLPNNKMLTTPESAHPTGVPPLKARSDIGSNVQYTNPNLHFI 179

Db 180 INVHDNCGNYSVLDKVLDRREVEPELRLTGLDGGSPRSGLTLLIITLINDVY 239  
180 VVHSRTDGRKPYELVLDRLADREGEPELTLITLALDGGAPSSGTTTIVIEVVDINDS 239

Db 240 PEYVESLYKVOVPENSVGLVYTVSARDIDTGSNGEIVYAFYATERTLKTFRINSTG 299  
240 POFVOSLYKVOVPENNLNFAVTVYSATIDAGYGNVTVSLPG-VGVQPPVIDEITG 298

Db 300 NLHKAEALNTOTYTLITQADGGGLSGKCTVYVAVTDINDNPPLMSSITSPEN 359  
300 EHLKSKLEDEEISNHNIEIAADTGGGLSGKCTVAVQVLDVNDNAPLELITRKLTVLPPEN 358

Db 360 SPETVAVFRIIRDSDGNNAKWCISIDHLFVLKPSVENFTYTLVTERALDREPEREYNI 419

Db 359 SAETVAVFVSVDSDSGDNBWCISIPNNIPFLKTFENYITVIEGDLDENAEYNI 418  
420 TITVTLGTPRLKTOGNLTVYSDVDNAPTISQTYTLRVKNNSPALHISVSATDRD 479  
419 TITVSDIGTPRLTTOHTITVOVSIDINDNAPAFQISYTFVHNNSPALHICTISATSD 478

Db 480 SGANAQVYSLPPEHDPOLPLGSLVSIINDGOLFALRSIDPEALQAFEFVGAADRGSP 539  
479 SGNAHTTISLPPDPOLALDSIINDGOLFALRDYDALQSFEFYVGTADGSP 538

Db 540 ALSSQALRYVLADANDNAPVLYPLQNGSAPCTELVPRAAGYLVAKVAVDSDGON 599  
539 ALSSQTLRYVVLDDNDNAPVLYPLQNSAPCTELLPRAABGYLITVAVDSDGON 598

Db 600 AWLSYOLKATEPGLFSVAHNGEVRT 626  
599 AWLSFOLKATEPGLFSVAHNGEVRT 625

RESULT 9  
US-09-099-639-112  
Sequence 112 Application US/09099639  
Patent No. 6262237  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,639  
FILING DATE: 18 JUN 1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/263,161  
FILING DATE: 27 JUN 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Greta E. No. 6262237and  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/34703  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 112:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 797 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-099-639-112

Query Match 62.6%; Score 2009.5; DB 4; Length 797;  
Best Local Similarity 62.7%; Pred. No. 4.5e-181;  
Matches 393; Conservative 83; Mismatches 148; Indels 3; Gaps 3;

Db 1 MEARVVALORQVSLVFLVGSW-AGAEPRLRYFAETERGFLNLAIDGLGYEEL 59  
1 METALAKIPOOROVFLTLT-LSLWKSSSEAIRSMPEETESGYMANLAKDGLRIGEL 59

Db 60 SARCGRIYSDETIGFLNPLTGLDLNLEKIDREELCGPTEPCVLPOLLERKFOIFRA 119  
60 SARCGRIYSDETIGFLNPLTGLDLNLEKIDREELCGPTEPCVLPOLLERKFOIFRA 119

```

Db 60 SSRGAQIHYKGNKELLQDAETGNLFLEKIDRELLGCEPVCVLIENPMQEFOT 119
QY 120 ELWVNDINDSHSPVFLDREITLTIESTPGATFLESADSDVGINNRYNTISSNYEH 179
Db 120 ELQITDINDHSEFPNKKMLLTIPESAHPTVPLKARDSDISNAQNTVPMNLFH 179
QY 180 INVHNGGNGVSELYLVDKVLDRREVPRLTLGLDGGSPRSCTYLIRLVLINDNV 239
Db 180 VVTHSRGTGRKPELVLDRLADREOPRLTLTALDGGAPRSCTYVHLEVDINDNS 239
QY 240 PEFVESLYKVOVPENSPVGLVTVSARDLDTGSGNGEIVAFYVATERLTKEFRINSTG 299
Db 240 PQFVQSLYKVOVPENNPFLNAFVTVSATDLGAGYGVNTYSLFQG-YGVFPQFVDEITG 298
QY 300 NHTLAELNYEAITOTYTLTIAOKDGGSLGKCTVYVHTDINDNPELMSLSPREN 359
Db 299 EHLKSKELDEFEISNHNIEATDGGISGKCTAVOVLDVNDNAPELTIKRLIVLVEN 358
QY 360 SPETVAVAFRIRDRDSGNNAKWCSDIHLPEVLKPSVENEFTYVTERALDREERTENI 419
Db 359 SAETVAVAFVSVDSDSGDNGMVCISIPNNIPFLKPTFENYTYLTGEPDLRENAEYNI 418
QY 420 TTVVDTGLPRLKTOHNLTVTVSDVNDNAPFTSOTTYTLRRENSPALHIGSVATRD 479
Db 419 TTVVDTGLPRLKTOHNLTVTVSDVNDNAPFTSOTTYTLRRENSPALHIGSVATRD 478
QY 480 SGANNOVTSLPPHDPOLPLGSLVSIADNGOLFALRSLDLEALQAEFRVGAADRGSP 539
Db 479 SGSMNHITYSLPPDDPOLADSLISINVDNGOLFALRSLDLEALQAEFRVGAADRGSP 538
QY 540 ALSSQALVRLVADANDNAPFLVPLONGSAPCTEIVPRAAEAGYLAKVAVADGSGON 599
Db 539 ALSSQTLVRLVADNDNAPFLVPLONGSAPCTEIVPRAAEAGYLITKVVAVADRDGON 598
QY 600 AWLSYOLKATEPGLFGVMAHNGEVRT 626
Db 599 AWLSFOLLKATEPGLFSVMAHNGEVRT 625

```

## RESULT 10

PCT-US95-08071-112  
Sequence 112, Application PC/TUS9508071

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 115

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

STREET: 6300 Sears Tower, 233 S. Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/08071

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/12588

FILING DATE: 23 DEC 1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/998,003

FILING DATE: 29 DEC 1992

ATTORNEY/AGENT INFORMATION:

NAME: Noland, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 32149

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ. ID NO: 112:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-08071-112

```

Query Match 62.6%; Score 2009.5; DB 5; Length 797;  
Best Local Similarity 62.7%; Pred. No. 4.5e-181;  
Matches 393; Conservative 83; Mismatches 148; Indels 3; Gaps 3;

```

QY 1 MEARVNAIAKROVSLICVFLGWSW-AGAEPLVFAVEEERETFLANLADIGLVEEL 59
Db 1 METALAKIPQOROVFFLT-LSLIMKSSSEAIRISMEETESGYMANLAKDGIRYGL 59
QY 60 SARCGRIVSDETIGFLINPLTGDILNEKIDREELGCTPEPCVLPFOLLERQFIFRA 119
Db 60 SSRGAQIHYKGNKELLQDAETGNLFLEKIDRELLGCEPVCVLIENPMQEFOT 119
QY 120 ELWVNDINDSHSPVFLDREITLTIESTPGATFLESADSDVGINNRYNTISSNYEH 179
Db 120 ELQITDINDHSEFPNKKMLLTIPESAHPTVPLKARDSDISNAQNTVPMNLFH 179
QY 180 INVHNGGNGVSELYLVDKVLDRREVPRLTLGLDGGSPRSCTYLIRLVLINDNV 239
Db 180 VVTHSRGTGRKPELVLDRLADREOPRLTLTALDGGAPRSCTYVHLEVDINDNS 239
QY 240 PEFVESLYKVOVPENSPVGLVTVSARDLDTGSGNGEIVAFYVATERLTKEFRINSTG 299
Db 240 PQFVQSLYKVOVPENNPFLNAFVTVSATDLGAGYGVNTYSLFQG-YGVFPQFVDEITG 298
QY 300 NHTLAELNYEAITOTYTLTIAOKDGGSLGKCTVYVHTDINDNPELMSLSPREN 359
Db 299 EHLKSKELDEFEISNHNIEATDGGISGKCTAVOVLDVNDNAPELTIKRLIVLVEN 358
QY 360 SPETVAVAFRIRDRDSGNNAKWCSDIHLPEVLKPSVENEFTYVTERALDREERTENI 419
Db 359 SAETVAVAFVSVDSDSGDNGMVCISIPNNIPFLKPTFENYTYLTGEPDLRENAEYNI 418
QY 420 TTVVDTGLPRLKTOHNLTVTVSDVNDNAPFTSOTTYTLRRENSPALHIGSVATRD 479
Db 419 TTVVDTGLPRLKTOHNLTVTVSDVNDNAPFTSOTTYTLRRENSPALHIGSVATRD 478
QY 480 SGANNOVTSLPPHDPOLPLGSLVSIADNGOLFALRSLDLEALQAEFRVGAADRGSP 539
Db 479 SGSMNHITYSLPPDDPOLADSLISINVDNGOLFALRSLDLEALQAEFRVGAADRGSP 538
QY 540 ALSSQALVRLVADANDNAPFLVPLONGSAPCTEIVPRAAEAGYLAKVAVADGSGON 599
Db 539 ALSSQTLVRLVADNDNAPFLVPLONGSAPCTEIVPRAAEAGYLITKVVAVADRDGON 598
QY 600 AWLSYOLKATEPGLFGVMAHNGEVRT 626
Db 599 AWLSFOLLKATEPGLFSVMAHNGEVRT 625

```

## RESULT 11

US-07-998-003A-107

Sequence 107, Application US/07998003A

Patent No. 5643781

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

STREET: 20 South Clark Street

```

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/998,003A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 564378land, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 30903
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/246-5750
TELEFAX: 312/984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 682 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-998-003A-107

Query Match          37.9%; Score 1217; DB 1; Length 682;
Best Local Similarity 41.0%; Pred. No. 3.3e-106;
Matches 254; Conservative 117; Mismatches 242; Indels 6; Gaps 4;

OY 12 ROYSLICVFLGVSNAGAPLRYFAVEETRGFTLANLAIDLGVEELSGRCRIVSDET 71
D 15 RVGVLT-LLGLANKASTVHYEIPEREKGFAGVANNVNLGLDGLSLARRPVSAGS 73
OY 72 IGFLLNPLTGDLINERLDEELCGPTEPCVLPFQILLEPQIFRAELMVRDINDHSP 131
D 74 RRFVEVNETGEMVNDRLDEELCGLPSCVTYLLVENVPLELFSVEVYIODINDNP 133
OY 132 VFLDREITNLLESTPGATFLESADSDVGINNLRNTISSNVYEHINVDNGECNVY 191
D 134 APTQEMKLEISEAVAGTRPLESHADPDGNSLQTYELSNVEFALRVOTREDSTKY 193
OY 192 SELVDKYLDEEYVELRLTLTGDDGSPRSSTLLRLVLDINDVPEVESLYKVQY 251
D 194 AELVLERALDEREPISQIYLTALDGTPLASLPHIKVLDANDNAPVFNOSLYRARV 253
OY 252 PENSPPGSLVVTYASARDLDTSGNGEIVYAF-FYATERLTKTRFNTSGNLHLKAEINYE 310
D 254 PGCGTSGTRVYQVATLDDEGPNGEIITISFGSHNRAGVRLPALDVTGLTKGRDPE 313
OY 311 AIQTYTLTQAKDGG--LSGCTVVVHYTDINDNPPELLMSSLTSPSPSPETVYAVF 368
D 314 DTKLHEIYIOAKDKGANPEGACHKVLVEVDVNDNAPETIVTSVSPVEDASGVYIAL 373
OY 369 RIRDROGNNAKMVCISIOHLPFVLPKPSVENFTLVTERALDREERTYNTITYVDLGT 428
D 374 SVYDLDAGEGLVTCVPPGLPSSLTSLKNFTYLTSTADLREYVPEVNSITARDAGT 433
OY 429 PRKTOHNTLVYSDVNDNAPFTSQTYYTLRRENNSPALHIGSVASDROGNAOVYX 488
D 434 PSLSALTIVRYOVSNDINPPOSSOSIYVIEENNLPAPIILNLSVWPDPAQNARLSF 493
OY 489 SLIPDPDPLIGSLVSNADNGQLEFALSLDEALQAEFRVGAADRSPALSSQALVR 548
D 494 FLLEQGAETGLVGRYFTINRNGIYSLVPLDYEDRREBELFAHISDGTVPYLAINISN 553
OY 549 VIYADANDNAPFVLYPLONGSAPCTELVYRAEAGTAVKAVVAVDGDSONAMLSYQLIK 608
D 554 IFVTDNRDNAPVLYPRPGSS--VEMLPKGTSGHLSRVVGVWDADAGHNAMLSTSLFG 611

```

```

OY 609 ATEPGLFGVWANGGEVETA 627
D 612 SPNQSILFAIGLHTGQISTA 630

RESULT 12
US-08-453-274B-107
Sequence 107, Application US/08453274B
Patent No. 5663300
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,274B
FILING DATE: 30-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: No. 5663300and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 32660
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 682 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-453-274B-107

Query Match          37.9%; Score 1217; DB 1; Length 682;
Best Local Similarity 41.0%; Pred. No. 3.3e-106;
Matches 254; Conservative 117; Mismatches 242; Indels 6; Gaps 4;

OY 12 ROYSLICVFLGVSNAGAPLRYFAVEETRGFTLANLAIDLGVEELSGRCRIVSDET 71
D 15 RVGVLT-LLGLANKASTVHYEIPEREKGFAGVANNVNLGLDGLSLARRPVSAGS 73
OY 72 IGFLLNPLTGDLINERLDEELCGPTEPCVLPFQILLEPQIFRAELMVRDINDHSP 131
D 74 RRFVEVNETGEMVNDRLDEELCGLPSCVTYLLVENVPLELFSVEVYIODINDNP 133
OY 132 VFLDREITNLLESTPGATFLESADSDVGINNLRNTISSNVYEHINVDNGECNVY 191
D 134 APTQEMKLEISEAVAGTRPLESHADPDGNSLQTYELSNVEFALRVOTREDSTKY 193
OY 192 SELVDKYLDEEYVELRLTLTGDDGSPRSSTLLRLVLDINDVPEVESLYKVQY 251
D 194 AELVLERALDEREPISQIYLTALDGTPLASLPHIKVLDANDNAPVFNOSLYRARV 253
OY 252 PENSPPGSLVVTYASARDLDTSGNGEIVYAF-FYATERLTKTRFNTSGNLHLKAEINYE 310
D 254 PGCGTSGTRVYQVATLDDEGPNGEIITISFGSHNRAGVRLPALDVTGLTKGRDPE 313
OY 311 AIQTYTLTQAKDGG--LSGCTVVVHYTDINDNPPELLMSSLTSPSPSPETVYAVF 368
D 314 DTKLHEIYIOAKDKGANPEGACHKVLVEVDVNDNAPETIVTSVSPVEDASGVYIAL 373
OY 369 RIRDROGNNAKMVCISIOHLPFVLPKPSVENFTLVTERALDREERTYNTITYVDLGT 428

```

```

Db 374 SVTDLGAGENGVLTCCEVPPGLPFLSTSLKNTYFLTKTSADLDRRETVENMLSTIARDAGT 433
QY 429 PRKTOHNLTVTVSDVNDNAAPTEFSQTYTLRVRENNSPALHIGSVKATDDSGANAOVY 488
Db 434 PSLTALTIVRVQVSDINDNPQSSOSDYVIEENNLPGAPILNLSWDDAPQNALSLF 493
QY 489 SLPPHDPOLPLGSLVSVINADNGQLFALRSLDPEALQAFEFVGAADRGSPALSSQALVR 548
Db 494 FLLEOGAETGLVGRYFTINRDNGLVSLPLDYDRREFELTAHISDGTPLVATINSVN 553
QY 549 VLVADANDANPFLVLYPLONGSAPCTELVPRAAEGYLVAKVAVDGDSONAMLSTYOLK 608
Db 554 IFVTRDNDNAPQVLYXPRPGSS--VEMLPRTSGHGLVSRVYVGDADAGHNAWLSTSLFG 611
QY 609 ATEPGLGVMAHNGEVRTA 627
Db 612 SPNOSLFAIGLHTGOISTA 630

```

## RESULT 13

```

US-08-453-695A-107
; Sequence 107, Application US/08453695A
; Patent No. 5708143
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,695A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5708143and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32658
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 682 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-695A-107

```

```

Query Match 37.9%; Score 1217; DB 1; Length 682;
Best Local Similarity 41.0%; Pred. No. 3.3e-106;
Matches 254; Conservative 117; Mismatches 242; Indels 6; Gaps 4;

```

```

QY 12 ROVSLVCLVFLGVSWAGAPLRYFAVEFERGTFLANLAIDGLVEEISANGCRIVSDET 71
Db 15 RVVGVTL-LILGALKMSKSVIHEIPEEREKGFVAVGVANGLDGLSISARRFPVSGAS 73
QY 72 IGFLLNPLTGLDLNLEKLDREELCGPTEPCVLPFOLLLEKPFQFERELAVRDINDNSP 131
Db 74 RREFVNRGTGEMFVNRDLREELCGTLPSCVTILELVENPLELSEVAVIYDINDNPP 133

```

```

QY 132 VFELREITINLESTPGATFELLESADSDVGINNLRNTYSSNVYEHINWHDNGEVNY 191
Db 134 APFTQEMKLEISEAVAPGTRFPLESHADBDLSNSLQTYELSRNEYFALRVQTRDSTKY 193
QY 192 SELVLDKVIDREVEPELRLTLTGDOGSPRRSGTTLIRIVINDNNVPEFESLYXVOY 251
Db 194 AELVLERALDREREPELQVLTALDGGTPALASLPITHIKVIDANDNAVEFQSLYRKY 253
QY 252 PENSPPGSLVTVVVSARDLDTGSGEIVYAF-EYATERTLTPRINSTSGNLKAEIYVE 310
Db 254 PGCTGSTRVVOYLALDDEGRNGEITTFSGSHNAGVQVLFALDLVYGMILKRLDPE 313
QY 311 ALOTYTLTLOAKDGG--LSGCTVAVVHTDINDNPELLMSLSPSPENSPEVYAVE 368
Db 314 DFKLHEIYIOAKDKGANPGACGKVLVEVVDVNDNAPETTVSVSPVEDASGVYIAL 373
QY 369 RTRDRSGNNAKAVGSIQDHLPRVLKPSVENRTYTLTEALDREREVENITITVDGT 428
Db 374 SVTDLGAGENGVLTCCEVPPGLPFLSTSLKNTYFLTKTSADLDRRETVENMLSTIARDAGT 433
QY 429 PRKTOHNLTVTVSDVNDNAAPTEFSQTYTLRVRENNSPALHIGSVKATDDSGANAOVY 488
Db 434 PSLTALTIVRVQVSDINDNPQSSOSDYVIEENNLPGAPILNLSWDDAPQNALSLF 493
QY 489 SLPPHDPOLPLGSLVSVINADNGQLFALRSLDPEALQAFEFVGAADRGSPALSSQALVR 548
Db 494 FLLEOGAETGLVGRYFTINRDNGLVSLPLDYDRREFELTAHISDGTPLVATINSVN 553
QY 549 VLVADANDANPFLVLYPLONGSAPCTELVPRAAEGYLVAKVAVDGDSONAMLSTYOLK 608
Db 554 IFVTRDNDNAPQVLYXPRPGSS--VEMLPRTSGHGLVSRVYVGDADAGHNAWLSTSLFG 611
QY 609 ATEPGLGVMAHNGEVRTA 627
Db 612 SPNOSLFAIGLHTGOISTA 630

```

## RESULT 14

```

US-08-268-161A-107
; Sequence 107, Application US/08268161A
; Patent No. 5798224
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/268,161A
; FILING DATE: June 27, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,337
; REFERENCE/DOCKET NUMBER: 27866/32149
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 682 amino acids

```

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
us-08-268-161a-107

Query Match 37.9%; Score 1217; DB 1; Length 682;  
Best Local Similarity 41.0%; Pred. No. 3,3e-106;  
Matches 254; Conservative 117; Mismatches 242; Indels 6; Gaps 4;

```

QY 12 ROYSLICVFLGYSWAGAEPLRYFAEETRGTEFLANLAIDLGLVEELSGRCRIVSDET 71
DB 15 RVYGLV-LTLGALNKASVYIHEIPEERKGFAGVNVANGLDGLGSLASARFPVSGAS 73
QY 72 IGFLLNPLTGLDLNKKDLREELCGPTEPCVLPOLLERFQIFRAELWVRDINDHSP 131
DB 74 RREFVENRTGEMFVNDRLDRRELCGTLPSCVTLELVENPLELFSEVYIODINDNNP 133
QY 132 VFLDRITLNLLESTTPGATFLLESADSDVGINNLRYVTSNNYFIHNDGEGVY 191
DB 134 APTQEMKLEISEAVAPGTRPPLSADHPDGLGNSLQYELSRNDEYFALRQTRDSTKY 193
QY 192 SELVLDKVLDRREVELRLTLTGDLGSPPRSGLTLRLIVLDINDVPEVESLYKVOY 251
DB 194 AELVLERALDRERPSLQVLTLALDGGTPALSASLPIHIKVLANDANAPVNOGLYRVR 253
QY 252 PENSVPGLVYVVSARDLDTGSGEIVYAF-FYATERLKTRFINSNGNLHKAELNYE 310
DB 254 PGCGTSGTRVQVLTALDDEGPNCEIITFSGSHRAGVROLFAIDLVTGMITIKGRDLE 313
QY 311 AIQYITLTIOAKDGG--LSGCTVYVHVTDINDNPELLMSSLTSPIPENSEVYAVF 368
DB 314 DTKLHEIYIOAKDKGANEGACHKVLVEVVDVNDNAEIVTYSVSPEDASGTVAL 373
QY 369 RIRPDGSGNNAMKVCSDIHLPEVLPKPSVENEFTLVTERALDREREYNTITVTDLGT 428
DB 374 SVTDLDAENGVLVCEVPPGLPSLTSSLKNYFTLKTSADLDRETYEYNLSTARAGT 433
QY 429 PRLKTOHNLTVYSDVNDNAPTFSQTYTLRVRENNSPALHIGSVATDRDSCANAQVY 488
DB 434 PLSALITIVRQVSDINDNPOSSQSYDYIEENNLPGAPILNLSWMDADAPANARLSF 493
QY 489 SLPPHPDPLGSLVSIINADNGOLFALRSIDPEALQAEFFRVGAADRGSPALSSQALVR 548
DB 494 FLBQGAETGLVGRYFTINRNGIVSSLVPLDYEDRREFELTAHISGCTPVATNINSV 553
QY 549 VLVADANDNAPFVLYPLONGSAPCTELVPRAEAGYLVAVVAVDDSGONAMLSYQLK 608
DB 554 IFVTDNRDNDAPVLYPRPGSS--VEMLPRTSAGHLVSRVVGMDADAGHNAWLSTSLFG 611
QY 609 ATEPGLFGVMAHNGEVRTA 627
DB 612 SPNOSLFAIGLHTGOISTA 630

```

## RESULT 15

US-08-453-702A-107  
Sequence 107, Application US/08453702A  
Patent No. 5891706

GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,702A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5891706and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32657  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ. ID NO: 107:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 682 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
us-08-453-702A-107

Query Match 37.9%; Score 1217; DB 2; Length 682;  
Best Local Similarity 41.0%; Pred. No. 3,3e-106;  
Matches 254; Conservative 117; Mismatches 242; Indels 6; Gaps 4;

```

QY 12 ROYSLICVFLGYSWAGAEPLRYFAEETRGTEFLANLAIDLGLVEELSGRCRIVSDET 71
DB 15 RVYGLV-LTLGALNKASVYIHEIPEERKGFAGVNVANGLDGLGSLASARFPVSGAS 73
QY 72 IGFLLNPLTGLDLNKKDLREELCGPTEPCVLPOLLERFQIFRAELWVRDINDHSP 131
DB 74 RREFVENRTGEMFVNDRLDRRELCGTLPSCVTLELVENPLELFSEVYIODINDNNP 133
QY 132 VFLDRITLNLLESTTPGATFLLESADSDVGINNLRYVTSNNYFIHNDGEGVY 191
DB 194 AELVLERALDRERPSLQVLTLALDGGTPALSASLPIHIKVLANDANAPVNOGLYRVR 253
QY 252 PENSVPGLVYVVSARDLDTGSGEIVYAF-FYATERLKTRFINSNGNLHKAELNYE 310
DB 254 PGCGTSGTRVQVLTALDDEGPNCEIITFSGSHRAGVROLFAIDLVTGMITIKGRDLE 313
QY 311 AIQYITLTIOAKDGG--LSGCTVYVHVTDINDNPELLMSSLTSPIPENSEVYAVF 368
DB 314 DTKLHEIYIOAKDKGANEGACHKVLVEVVDVNDNAEIVTYSVSPEDASGTVAL 373
QY 369 RIRPDGSGNNAMKVCSDIHLPEVLPKPSVENEFTLVTERALDREREYNTITVTDLGT 428
DB 374 SVTDLDAENGVLVCEVPPGLPSLTSSLKNYFTLKTSADLDRETYEYNLSTARAGT 433
QY 429 PRLKTOHNLTVYSDVNDNAPTFSQTYTLRVRENNSPALHIGSVATDRDSCANAQVY 488
DB 434 PLSALITIVRQVSDINDNPOSSQSYDYIEENNLPGAPILNLSWMDADAPANARLSF 493
QY 489 SLPPHPDPLGSLVSIINADNGOLFALRSIDPEALQAEFFRVGAADRGSPALSSQALVR 548
DB 494 FLBQGAETGLVGRYFTINRNGIVSSLVPLDYEDRREFELTAHISGCTPVATNINSV 553
QY 549 VLVADANDNAPFVLYPLONGSAPCTELVPRAEAGYLVAVVAVDDSGONAMLSYQLK 608
DB 554 IFVTDNRDNDAPVLYPRPGSS--VEMLPRTSAGHLVSRVVGMDADAGHNAWLSTSLFG 611
QY 609 ATEPGLFGVMAHNGEVRTA 627
DB 612 SPNOSLFAIGLHTGOISTA 630

```

Search completed: May 16, 2003, 10:49:15  
Job time : 20 secs

Fri May 16 11:36:27 2003

us-10-050-704-125.ra1

Page 11





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 10:46:02 ; Search time 21 seconds  
(without alignments)  
2870.299 million cell updates/sec

Title: US-10-050-704-125

Perfect score: 3208

Sequence: 1 MEARVVALQKRVSLCVF.....KATEPGLFGVWANGGEVTA 627

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1989.5	62.0	701	T17243	hypothetical prote
2	1094	34.1	889	T09055	protocadherin 68 -
3	900.5	28.1	1180	T31066	vascular cadherin-
4	807	25.2	1072	T00041	BH-protocadherin p
5	799	24.9	1069	T00043	BH-protocadherin-a
6	697.5	21.7	1200	T00042	BH-protocadherin p
7	596	18.6	5147	T14119	cadherin-related t
8	584	18.2	3034	T14119	seven-pass transme
9	543.5	16.9	2610	T20968	hypothetical prote
10	542	16.9	4351	T00252	MEGFI protein - ra
11	485	15.1	2163	T15276	hypothetical prote
12	453	14.1	913	A47543	R-cadherin precurs
13	445.5	13.9	916	C38992	cadherin 4 precurs
14	442.5	13.8	3343	S44887	ZK112.7 protein -
15	438	13.7	790	G02678	cadherin-14 - huma
16	437.5	13.6	785	I50180	cadherin-7 - chick
17	435	13.6	913	IUCHCR	R-cadherin precurs
18	435	13.6	3097	T00021	DN-cadherin - frui
19	430.5	13.4	789	I52701	K-cadherin - rat
20	430.5	13.4	4307	T20721	cadherin-6B - chic
21	429	13.4	4307	T20721	hypothetical prote
22	428	13.3	1439	T27110	hypothetical prote
23	426.5	13.3	790	I37016	cadherin-6 - huma
24	417.5	13.0	794	I51638	F-cadherin - Afric
25	413.5	12.9	794	I59372	cadherin 12 - huma
26	403.5	12.6	905	D38992	cadherin 8 - huma
27	403.5	12.6	905	IJALC1	N-cadherin 1 precu
28	400.5	12.5	906	IJALC2	N-cadherin 2 precu
29	396	12.3	620	T15273	hypothetical prote

30	395	12.3	2809	2	T30213	G-cadherin - sea u
31	390	12.2	877	1	IJBOCN	N-cadherin precurs
32	386.5	12.0	906	1	IJMSCN	N-cadherin precurs
33	384.5	12.0	796	2	I48277	cadherin-11 - mous
34	383	11.9	709	2	T46413	hypothetical prote
35	382	11.9	796	2	I49556	cadherin-11 - mous
36	382	11.9	906	1	IJHUCN	cadherin 2 precurs
37	380.5	11.9	796	2	A53584	OB-cadherin precurs
38	380.5	11.9	871	2	S47518	cadherin - African
39	378	11.7	796	2	A38992	cadherin 11 precur
40	374	11.8	912	1	IUCHCN	N-cadherin precurs
41	361.5	11.3	784	1	IJHUC5	cadherin 5 precurs
42	360.5	11.2	373	2	T34563	hypothetical prote
43	354.5	11.1	887	1	IUCHCL	E-cadherin precurs
44	353	11.0	829	1	IJHCHP	cadherin 3 precurs
45	351.5	11.0	783	2	I50116	N-cadherin precurs

#### ALIGNMENTS

##### RESULT 1

T17243  
hypothetical protein DKFZp586B0217.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T17243

R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, September 1999

A:Reference number: Z18722

A:Accession: T17243

A>Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-701 <KOE>

A:Cross-references: EMBL:AL117449

A:Experimental source: adult uterus; clone DKFZp586B0217

C:Genetics:

A>Note: DKFZp586B0217.1

Query Match	Best Local Similarity	62.0%	Score 1989.5	DB 2: Length 701:
Matches 381; Conservative 58; Mismatches 93; Indels 1; Gaps 1;				
QY 95	LCGPTEPCVLPFOELLEPEQIFRAELVWRINDSHSVFLDRITLTLSTPGATFLL 154			
DB 1	MCGATEPCILHFOLELLENPVQFOTDQLDINDINHAEPFKEMLTIPESTQGTVPFL 60-			
QY 155	ESAHSDVGINNLNRYTSSNVYFHINVDNGEGNVYSELVLDKVDREVPETRLTLTG 214			
DB 61	KIADPDFIGSNVTYNTISPSHFHVATHNKGDRKYPDELVDKALDRERPELSTLTA 120			
QY 215	LDGSPRPSGTTLRILVLDINDONVPEFVSLYVQVSPVSLVYVSARDLDTGSN 274			
DB 121	LDGAPPSPSGTTLRILVLDINDONVPEFVSLYVQVSPVSLVYVSARDLDTGSN 180			
QY 275	GEIYVAFYATERTLTKFRINSTGNLTKAELNVEAIQYVTLTIQAKDGSGLSGKTVV 334			
DB 181	GSVAVAFLEQSDPELT-QPVIDEKTAETRLKRAKALDFEATPYINVAITVTDGGLSGKCTVA 239			
QY 335	VHVTDINDNPPELLMSITSPISPENSPETVYAVFRIRDRDSGNNAKAVCSIQHLPEVLK 394			
DB 240	IEVVDVNDNAPELMTSTSPENAPETVYAVSVSDPSGDGMRMICSIONLPEFLK 299			
QY 395	PSVNEFTVLYERLADREPEREYMITTYMDTIGPRLKTQHNLVYVSDVNDNAPPEFSOT 454			
DB 300	PTLNFETVLYQRLDRSQAEYMITTYMDTIGPRLKTQHNLVYVSDVNDNAPPEFSOT 359			
QY 455	TYTLRVRENNSPALHIGSVSATDRSGANQVYSLPDPOLPLGSLVSIANDGOLF 514			
DB 360	SYTLFVRNNSPALHIGSVSATDRSGANQVYSLPDPOLPLGSLVSIANDGOLF 419			
QY 515	ALRSIDFPAIDAEFRRYGAADRGSPLASQALVYLVADANDAPVLYLQNGAPCTE 574			

Db 420 ALRSLDYLALQAFERRVATDRGSPALSSSEALVRLVLDANDNSPVLVPLONGSAPCTE 479  
 QY 575 LVPRAAAGYLVAKVAVADGSGONAMLSYOLLKATEBGLGVVAHNGEVRTA 627  
 Db 480 LVPRAAAGYLVTKVAVADGSGONAMLSYOLLKATEBGLGVMAHNGEVRTA 532

## RESULT 2

T09055  
 Protocadherin 68 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
 C:Accession: T09055  
 R:Jin, P.; Xu, H.; Israel, D.  
 Submitted to the EMBL Data Library, October 1997  
 A:Reference number: 216540  
 A:Accession: T09055  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-889 <JIN>  
 A:Cross-references: EMBL:AF029343; NID:g2599501; PID:g2599502  
 C:Genetics:  
 A:Gene: PCH68

Query Match 34.1%; Score 1094; DB 2; Length 889;  
 Best Local Similarity 38.5%; Pred. No. 1.3e-64;  
 Matches 252; Conservative 116; Mismatches 235; Indels 52; Gaps 16;

QY 14 VSLLCVFLVSWAGA---EPLRYFAEETRGTFPLANLALDIDGLVEELSGARCVSDRTIGFLL-----NPLTG 82  
 Db 3 LSIICCCFL--LWAPALTKLNLNTSVPEEGAGVYIGNIGDARQGLPPLAENGCGGRSK 60  
 QY 64 ---CRIYSDETIGFLINPLTGLDLLNEKDLRELCGPTPCVLPFOULL-EXPFOIFRA 119  
 Db 61 SGSTRVLENSAPRLLDVADDSGLLYTKORIDRESLCRNNAKCOLSLTFVANDKEICMIKV 120  
 QY 120 ELAWNDIDHSPVFLDREITLNTLESTTPGATFLESASHPDVGNNLRVYLTSSNVY-- 177  
 Db 121 E--TQDINDNAPRSSQIYXIDISENAPGTRPPLTSABPDGEGNCLRTYLLRPDQGL 178  
 QY 178 FHINVDHNGEGNVYSELVLDKVDREVPBELRLTLTGIDGSPRSSTGLIRILVLDIND 237  
 Db 179 FGLVYKSRGDDGTKEPPELVYQKALDRQNNHTLVLTALDGGEPFRSATVGINVKVIDSND 238  
 QY 238 NVPEFVESLYKVQVPEPSPGSLVYVSARDLTGSGEIVYAF-FYATERTLKTIRINS 296  
 Db 239 NSPVEAPSYIVELPEPNTPLGTVIDLNATDADGEPNGEVLVSFSVVPDRVRELEFSIDP 298  
 QY 297 TSGMLHKAELNYEAIQTYTLTIQAKDGGG--LSGCTVYVHYTDINDNPPEL-----LM 349  
 Db 299 KTGLIRKGNLDYENGMLEIDVQARLGLPRLIPAHCKVTVKLLDRNDNAPSIGEVSVRO 358  
 QY 350 SLSLTPSPENSPETVAVAFRIIDRDSGNNAKWCST-----QDHLPEVLK 394  
 Db 359 GALSSEAP---PGTVIALVRYTDRDSGKNGLOCRVLGGGGVGGGGVGGPGSGVFFKLE 415  
 QY 395 PSVNEFTVYTERALDREERTENITTYVDLCTPRKLTQHNLTIVYSDVNDNAPFSSQT 454  
 Db 416 ENDNFTVYVDRMDERTDEYNTVITVARDGSGPPLNSTKSPAIKILDNDNDPFRFKG 475  
 QY 455 TYTLRRENNSPALHIGSVSATDRDSCANAOVYTSLLPPDPQLPGSLVSTINAADGOLF 514  
 Db 476 LTVYQVHNNIPGDYKLSVLAQDPDGLQNGTSTSLPSHIGVSTIYTVSVNPTGATLY 535  
 QY 515 ALRSLDEALQAEFRVGAADRGSPA--LSSQALVRYLVADANDNAPFVLYP--LQNGSAPC 572  
 Db 536 ALRSENFEGTKAEFEKVLANDSGAPAHLESNATVRYTVLDVNNNAVIVLPTIQNDTA-- 593  
 QY 573 TEL-VPPAAAGYLVAKVAVADGSGONAMLSYOLLKATEBGLGVVAHNGEVRT 626  
 Db 594 -ELQVPRNAGIYLVSTVRLADSDDFGSGRLTEIYDGNDDHLFEIDPSSGEIRT 647

## RESULT 3

T31066  
 vascular cadherin-2 - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000  
 C:Accession: T31066  
 R:Tejo, P.; Breviarlo, F.; Huber, P.; Panzerl, C.; Dejana, E.  
 J. Biol. Chem. 273, 17565-17572, 1998  
 A:Title: Identification of a novel cadherin (vascular endothelial cadherin-2) located  
 A:Reference number: 220962; MUID:98316522; PMID:9651350  
 A:Accession: T31066  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1180 <TEJ>  
 A:Cross-references: EMBL:Y08715; NID:g2764993; PIDN:CAA69965.1; PID:g2764994  
 A:Experimental source: brain capillary

Query Match 28.1%; Score 900.5; DB 2; Length 1180;  
 Best Local Similarity 34.7%; Pred. No. 1.4e-51;  
 Matches 218; Conservative 114; Mismatches 241; Indels 55; Gaps 11;

QY 31 LRYFAEETRGTFPLANLALDIDGLVEELSGARCVSDRTIGFLL-----NPLTG 82  
 Db 32 VKFOYTEVPSTGTVIGKLS-----QELRVERKRGKAGDAFQILLQALPVMNSDGG 84  
 QY 83 DLLNEKDLRELCGPTPCVLPFOULLKEPQIFRAELWVDINDHSPVFLDREITLNI 142  
 Db 85 LSTSSRLDRKLCQEDPCLVSFDVLTAGASALHVEIYQVLDINDHQPOEPKQDELEI 144  
 QY 143 LESTPGATFLESASHPDVGNNLRNTISSNVFHHNVHDNGGCVNSSELYLDKVIDR 202  
 Db 145 SESASLHTRIPIDRLADDTGPNLSYSLSPSEHFALDYVGPDETHAEVYVKKELDR 204  
 QY 203 EEPPELRLTLTGIDGSPRSSTGLIRILVLDINDNVEPEFESLYKVQVPEPSPGSLVY 262  
 Db 205 ELHSYFDVLTAYKNDGNPPKSGISVYKVVNDNSNSVFAESSIALTEIPEDYPTGLILI 264  
 QY 263 TVSARDLTGSGEIVYAF-FYATERTLKTIRINSTGMLHKAELNYEAIQTYTLTQIA 321  
 Db 265 NLTATDPDQPGNGEVEFEFGKHVSPEVMNTFGIDAKTQIILLQALDYEKRPAYEVQYA 324  
 QY 322 KDGG--GLSGCTVYVHYTDINDNPPEL--SLSLTPSPENSP-ETVYVAFRIIDRDSG 376  
 Db 325 RDGPNISIPGHCKVLTKVDVNDNAPSLITWASQTSLSVEDLPDSFSLVSAVANDSDSG 384  
 QY 377 NNAKWCSTIODHL-PEVLKPSVENFTYTERALDREERTENITTYVDLCTPRKLTQH 435  
 Db 385 NNGLVHCHMLNQGELGHFRKRTNGNTYMLTNATIDREQMPYITLVFQDDGPPPLSAEK 444  
 QY 436 NLTIVSDVNDNAPFSSQTYTLRVRNNSPALHIGSVSATDRDSCANAOVYTSLLPPHD 495  
 Db 445 ELQIOVSDVNDNAPVEFKSRYEVSFWENNPPSLHILITLKAHDADGSGKGYRIT----- 499  
 QY 496 POLPGSLVSTINAADGOLFALRSIDFEALQAFEPVYGAADGSPRLSSQALVRYLVADAN 555  
 Db 500 KDSVSHLVITDFETGEVYTAQSLDYEQWAGFERQVLAEDRGQPOLASSISVWVSLDAN 559  
 QY 556 DNAPFVLYP-LQNGSAPCTELY-----PRAEA--GYLYA 587  
 Db 560 DNAPFVLYPVLSEKATISLVYNASTGHLLPIENPSCMDRAGTGIIPKKAHSPWSTLL 619  
 QY 588 KVVAVDGSOGONAMLSYOLLKATEBGLF 615  
 Db 620 TIVARDADSGANGELFYSIGDNDHLEF 647

## RESULT 4

T00041  
 Bh-protocadherin PCDH7 (clone Bh-Pcdh-b) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 01-Dec-2000  
 C:Accession: T00041; T00040  
 R:Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.; Sugano, S.

Genomics 49, 458-461, 1998  
A:Title: Cloning, expression analysis, and chromosomal localization of BH-protocadherin  
A:Reference number: Z14074; MUID:96277460; PMID:9615233  
A:Accession: T00041  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1072 <YOS>  
A:Cross-references: EMBL:AB006756; NID:g2579419; PIDN:BA025195.1; PID:g2579420  
A:Experimental source: clone BH-Pcdh-b  
A:Accession: T00040  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1058, 'LH', '1060', 'LH', '1063', 'Y', '1065', 'TVFG', <YO2>  
A:Cross-references: EMBL:AB006755; NID:g2979417; PIDN:BA025194.1; PID:g2579418  
A:Experimental source: clone BH-Pcdh-a  
C:Genetics:  
A:Map position: 4p15

```

A:Reference number: M41087; MUID:92069752; PMID:1959133
A:Accession: M41087
A:Molecule type: mRNA
A:Residues: 143-485;1279-5147 <MAH>
A:Cross-references: GB:M80537
A:Accession: B41087
A:Molecule type: DNA
A:Residues: 1-142;487-1278 <MA2>
A:Cross-references: GB:M80537
A>Note: 1229-Gly and 1233-Ser were also found
C:Genetics:
A:Gene: fat
A:Cross-references: FlyBase:FBgn0001075
C:Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homo
C:Keywords: calcium binding; cell adhesion; duplication; transmembrane protein
E:1-35/Domain: signal sequence #status predicted <SIG>
E:36-5147/Product: cadherin-related tumor suppressor #status predicted <MAT>
E:36-4583/Domain: extracellular #status predicted <EXT>
E:51-156/Domain: cadherin repeat homology <CR1>
E:159-270/Domain: cadherin repeat homology <CR2>
E:271-382/Domain: cadherin repeat homology <CR3>
E:390-494/Domain: cadherin repeat homology <CR4>
E:497-599/Domain: cadherin repeat homology <CR5>
E:602-708/Domain: cadherin repeat homology <CR6>
E:718-822/Domain: cadherin repeat homology <CR7>
E:831-942/Domain: cadherin repeat homology <CR8>
E:948-1049/Domain: cadherin repeat homology <CR9>
E:1052-1153/Domain: cadherin repeat homology <CR10>
E:1156-1278/Domain: cadherin repeat homology <CR11>
E:1281-1384/Domain: cadherin repeat homology <CR12>
E:1387-1489/Domain: cadherin repeat homology <CR13>
E:1492-1601/Domain: cadherin repeat homology <CR14>
E:1607-1713/Domain: cadherin repeat homology <CR15>
E:1717-1823/Domain: cadherin repeat homology <CR16>
E:1826-1922/Domain: cadherin repeat homology <CR17>
E:1925-2027/Domain: cadherin repeat homology <CR18>
E:2028-2167/Domain: cadherin repeat homology <CR19>
E:2169-2278/Domain: cadherin repeat homology <CR20>
E:2281-2384/Domain: cadherin repeat homology <CR21>
E:2387-2491/Domain: cadherin repeat homology <CR22>
E:2494-2596/Domain: cadherin repeat homology <CR23>
E:2599-2703/Domain: cadherin repeat homology <CR24>
E:2707-2810/Domain: cadherin repeat homology <CR25>
E:2813-2913/Domain: cadherin repeat homology <CR26>
E:2915-3013/Domain: cadherin repeat homology <CR27>
E:3014-3124/Domain: cadherin repeat homology <CR28>
E:3127-3229/Domain: cadherin repeat homology <CR29>
E:3232-3334/Domain: cadherin repeat homology <CR30>
E:3337-3439/Domain: cadherin repeat homology <CR31>
E:3442-3545/Domain: cadherin repeat homology <CR32>
E:3548-3651/Domain: cadherin repeat homology <CR33>
E:3654-3756/Domain: cadherin repeat homology <CR34>
E:3954-4010/Domain: EGF homology <EG1>
E:4017-4048/Domain: EGF homology <EG2>
E:4056-4089/Domain: EGF homology <EG3>
E:4096-4127/Domain: EGF homology <EG4>
E:4384-4609/Domain: transmembrane #status predicted <TM>
E:4610-5147/Domain: intracellular #status predicted <INT>

Query Match 18.6%; Score 596; DB 1; Length 5147;
Best Local Similarity 27.6%; Pred. No. 2,3e-30;
Matches 201; Conservative 98; Mismatches 258; Indels 172; Gaps 23;

QY 49 AIDGLVEEESA--RGGRVSDETIGFL-----LNLPLTGLLNEK 89
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 64 AVDSAPDPEVLEGPRG-----TTVGRTIPKPKFSYAFNPPREFLIDPTGGEKRVV 117

QY 90 LDREELCGPPEPCVLPFOLLLEKPFQIFRAELRWKDINDSPVLDREITINLESTPG 149
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 118 LDRGMDHDLVLLSSO-----PTYLEVRIKVLVDNDSPEPPEPDSIAISFESATSG 172

QY 150 ATFLLESHDSDVGINNLRN--TISNV--YFHINVDNDEGNV--SEVLVDKVLVDREE 204
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

Db 173 TRLLDAATDADYGVNGVTDQEIYVAGNVNKKFRLVTTANPSGDTSVLHLETTGNLDRES 232  
 QY 205 VVELLTLITGLDGGSPRPSGTLILILVDINDNPEVESEYKQVENSPPVSLVTV 264  
 Db 233 KGSYLINISARGGSPRGLQVAVNTILDVNDNPIDHDYVNLNETLPGTPVTV 292  
 QY 265 SARDDTSGNCEIVAFVAFATERTLKTFRIN-----STGNLHKALNEAI-----OTY 315  
 Db 293 MASDNDLGNKITY---YIAE-THQSTVNPETGISTEVNCKQTNKSSAQSC 348  
 QY 316 TLITIAKDGCG--LSGKCTVVVHTVDINDNPELIM-----SSLTSPLENSPE-TVAV 367  
 Db 349 VTFVARDHSGSPROGRTVYVNLDTNDHPIISFREPDGKVAIVDENAVNCTVAA 408  
 QY 368 FRIARDSDGNNAKWCST--QDHPFVLKPSVENVYTLVTEALDKERTENITITVD 425  
 Db 409 VAVKSDSGNLNRTSVRIVSGNELGHFRLERADLHIVRNGVLDKEIKGNLTVAMD 468  
 QY 426 LCPRLKTOHNLTVVSDVNDNAPTESOTYTLRVRENNSPALHIGSVASATRDGANAQ 485  
 Db 469 QSTPATTAAHLIIVNDVNDHEPFEKSEYSAVLSLAPGSEFVASITADEDTGVNAQ 528  
 QY 486 VVYSIL----- 491  
 Db 529 VHYDILSGNELKWFMSMDPLTGLITVTPLRERIDVELISARDGCPNPKFAYTQKYI 588  
 QY 492 -----PP-----HDPLPLGSLV----- 504  
 Db 589 IIDENDEAPQFSQREONVTLGEDAPPQITVALMTATDHD--QGTNCSVTFPALAPSVERTLP 647  
 QY 505 ---SINADNGQLFALRSLDFEALQAFEFVGAADRGSPA--LSSQALVRLVADANDANP- 559  
 Db 648 LQFALDALGQLTRPRDREKMSQEIPIYIARDGAPTPGATPVMINADVNDNPQ 707  
 QY 560 ----EVLVYLQNGSAPCTELVPRAAEGLYAKVAVVADSGSQNMLSTQILKATEPGLE 615  
 Db 708 FYPRIYISLTDDDDIK--LKEVEKERILHVTASDKDDGNALIEYRLESIGE-GLF 764  
 QY 616 GWAHNGEV 624  
 Db 765 QLDARSGAI 773

## RESULT 8

T1419  
 seven-pass transmembrane receptor protein precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T1419  
 R:Hadjantonakis, A.K.; Formstone, C.J.; Little, P.F.R.  
 submitted to the EMBL Data Library, October 1997  
 A:Description: The Celser family of novel evolutionarily conserved seven-pass transmembrane  
 A:Reference number: Z17881  
 A:Accession: T1419  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3034 <HMD>  
 A:Cross-references: EMBL:AF031572; NID:g3800735; PID:g300736; PIDN:AAC68836.1  
 C:Genetics:  
 A:Gene: Celser1  
 A:Map position: 15  
 C:Keywords: transmembrane protein  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:27-3034/Product: seven-pass transmembrane receptor protein #status predicted <MAT>

## Query Match

Best Local Similarity 18.2%; Score 584; DB 2; Length 3034;  
 Matches 180; Conservative 95; Mismatches 245; Indels 80; Gaps 17;

QY 66 IVSDERTIGLLNPLTGDLINKDLREEL-----CGTEFCVLPFQULLEKRPQ 115  
 Db 513 IVSGNLKGFYHLSGSDLVNPLDFEAIREYTLRIKADGGRPPLINSSGLV----- 566

QY 116 IFRAELVNRDINDHSPVFLDREITLNLLESTPGATFLLESADSDVGINNENYTI--- 172  
 Db 567 ----SVQVLDVNDNAPLFVSPFOAVLENVPLGHVSLHIGQADADAGENARQRYLVPI 622  
 QY 173 -----SSNVFHNVD--NCGNVSELYDKVLDREPELRLITGLD 216  
 Db 623 ASTIVGSSVDSSENPAPADPFPOIHNSGMITVCAE-----LDREVEHYSFGVEAVD 676  
 QY 217 CGSPRSGTLLIRILVLDINDNPEFVESLYKQVENSPPVSLVTVSARODTQSGNE 276  
 Db 677 HGSPAMSSASVSLTVLDVNDNPMFTQPYELRLNEDAAVSSVLTFLRRDRDANS--- 733  
 QY 277 IVYAFATERTLKTFRINSTGN--LHKALNEYALQTYTLTIAKDGGLSGKCTVV 334  
 Db 734 -VITYQLGCTNRFRALSSQSGGLTLLALPDYKQERYLAIVASD--GTSHSHAQVE 791  
 QY 335 VHYTDINDNPELIMSLTSPIPENSP-ETVAVFRIARDSDGNNAKWCSTODHLP-FY 382  
 Db 792 INVDTAMTHRPVQSSHYTVSVSEDRPVGTISATISATDEDTENARITVLEDVPVOPR 851  
 QY 393 LKPSVENFYLTVMERALDREERTENITITVDLGPRLKTOHNLTVVSDVNDNAPTS 452  
 Db 852 IDPDTGTYTM-TE--LDYEDQAYTLATAQDNGIPQKSDTYSLELILADANDNAPRL 908  
 QY 453 QTYTLRVRENNSPALHIGSVASATRDGANAQVYSILPDPHPLPLGSLVSINADNG 512  
 Db 909 RDTFGSVFEDAPRSTSVLQVSAATRDGSGNRLLYTFQGGDGD--GDFT-1EPTSGV 964  
 QY 513 LPAIRSLDFEALQAFEFVGAADRGSP-AISSQALVRLVADANDANP-----EVLVYL 565  
 Db 965 IRTORRLDRENVAAVYNLMALAVDRGSPNPLSASGIVSVLINDNPEVEKDELTEVE 1024  
 QY 566 QNGSAPCTELVPRAAEGLYAKVAVVADSGSQNMLSTQILKATEPGLEGVWAHNGEV 625  
 Db 1025 EN-----SPGVSVARIRANDPDGPNQAQIITVQIVEGVVEVQDLSSDLR 1072

## RESULT 9

T20968  
 hypothetical protein F15B9.7 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T20968; T26278  
 R:Percy, C.  
 submitted to the EMBL data library, August 1996  
 A:Reference number: Z19351  
 A:Accession: T20968  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2610 <MIL>  
 A:Cross-references: EMBL:Z78013; PIDN:CAB01427.1; GSPDB:GN00023; CESP:F15B9.7  
 A:Experimental source: clone F15B9  
 R:Baynes, C.  
 submitted to the EMBL Data Library, August 1996  
 A:Reference number: Z20187  
 A:Accession: T26278  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2610 <MIL>  
 A:Cross-references: EMBL:Z78013; PIDN:CAB01449.1; GSPDB:GN00023; CESP:F15B9.7  
 A:Experimental source: clone W07G4  
 A:Gene: CESP:F15B9.7  
 A:Map position: 5  
 A:Introns: 169/2; 217/3; 239/3; 284/3; 515/2; 550/3; 738/3; 810/1; 838/3; 890/2; 977/

## Query Match

Best Local Similarity 16.9%; Score 543.5; DB 2; Length 2610;  
 Matches 172; Conservative 100; Mismatches 271; Indels 115; Gaps 17;

QY 36 AEETRGFTLANLADLGLV-----EELSARG---CIVSDERTIGLLNPLTGDLIN 87  
 Db 376 ASDSYNVTILENTITPAVIAIVATKDEDEGTNGKAVHSMASSSGIGLITDYSTGEVTLR 435

QY 88 EKLD-----REELCGPTPCVLPFOLLLEKPFQIFRAELMWINDSHSPVFLDR 136  
 Db 436 ERIDAKNSPIYAVIRAKDAQPALSTVPLFIN-----VIDINDHAPFLILAA 482  
 QY 137 ETLNLITSTPGATFLLESASHDSVGINNLNRTISSNVFHNHNDGNGVNSVL 196  
 Db 483 QKMTLEENVAIGEENGVAVDAIDEDSGPNGLIKSMEGSEDFIID-EDSG-----LTKT 535  
 QY 197 DKVIDREVEPELRILTLTGIDGSPRSGTTLIRLVIDINDNVEFESEKTKVOPENSP 256  
 Db 536 TKLIDRETTAVSLKVTARMDGTPSLNNTIAVVLKINDINAPTEFKKEYNITISEMP 595  
 QY 257 VGSLLVYVSANDLDTGSGNGELIVAFYATERLTFTFRINSTGNLHLKALNEAIOITY 316  
 Db 596 RGSQITLTKA---VDNDEOKITRYRIEADREVFSIIDIGDGAITSVSGELKRQD-HKVR 652  
 QY 317 LTIOAKDGGGSGKCTVYVHTDI-----NDNPELLMSLTSPIDENSP-ETVV--- 365  
 Db 653 VEISTDGGIGQKCVAVNFIDVNSAPYENDHP-----FSVKIPEHSPIGYPIVITLK 705  
 QY 366 -----AVFIRIDRDSGNNAKMCISIODHLPEVLPKS 396  
 Db 706 VSFYFEGFYGLKHFYVSSNLKKRIPNSANMWAHDHRODNARIYVSISSQEFFRIDS 765  
 QY 397 VENFTYLTALRERTEYNITITVDTGTPRLKTQHNLVTYVSDVNDNAPTFSGTTY 456  
 Db 766 SGD---TSVSSDLDRERATEFSVTASDHASPLNTSTOIEVLDDINDNSPQFTSSY 822  
 QY 457 TLARVENSPPALHIGVSATDRDSCGANAQYTSLLPDPQLPGSLVSTINANGOLFAL 516  
 Db 823 AATISEDIPVGTSLQVSAIDADIGRNGIYDYFL-NESSSPSIQLFRIDRSTGLRVS 880  
 QY 517 RSLDFELQAFEEFVGAADRGSPALSSQALVRLVADANDNAP-----FVLYPLONGSA 570  
 Db 881 SKLDREFAVIVLPIFARDRGTSLSAASEITLILSDVNDNAPTEFQSLDYIAEN--- 937  
 QY 571 PCTELVPRAEAGYLVAVVADGSGONAMLSYOLLAKEPELFGV---MAINGEVR 625  
 Db 938 -----SPVGSVTGTIVARADBDGDNADISFRIFGADAKLFIEEDAONGVYR 986

## RESULT 10

T00252  
 MEGF1 protein - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: T00252  
 R:Nakajima, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.  
 Genomics 51, 27-34, 1998  
 A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs  
 A:Reference number: Z14126; MID:98360089; PMID:9693030  
 A:Accession: T00252  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-431 <NAK>  
 A:Cross-references: EMBL:AB011527; NID:93449285; PIDN:BA3458.1; PID:93449286  
 C:Genetic source: brain  
 A:Gene: MEGF1  
 C:Superfamily: rat MEGF1 protein; cadherin repeat homology; EGF homology; laminin G repeat  
 F:3798-3949/Domain: laminin G repeat homology <LGR>  
 F:3953-3985/Domain: EGF homology <EGF>  
 F:3992-4023/Domain: EGF homology <EGF1>

## Query Match

Best Local Similarity 16.9%; Score 542; DB 2; Length 4351;  
 Pred. No. 6.8e-27;  
 Matches 178; Conservative 99; Mismatches 238; Indels 160; Gaps 20;

QY 65 RIVSDEITGILLNPLTGOLLNEKIDRELCGPTECVLP-FOLLLEKPFQIF----- 117  
 Db 1803 KILPEALKFFKIDPSMGTLTTSSELDIED-----TFLFQ-----FNIYVHDGQT 1847

QY 118 -----RAELWVDINDHSPVELDREITNLITESTPGATFLLESASHDSVGINNR 168  
 Db 1848 PILFARSAKVIITHVDVNDSPPRESEQIYEVAVEPIHPGKGLTLVQAEKNDSDR- 1903  
 QY 169 NYTI-SSNVFHNVA-DNGEGNVYS-----ELVLDKVID 201  
 Db 1904 TYSIKTSNADAVTIIPTTGOISVNPATRLRFQKFSIRASDGLYHDAVVAISLTVLD 1963  
 QY 202 R-----EEPELRLT-----LTGLD----- 216  
 Db 1964 KSLQFDQDVAARAVENTPRKALVILGVGHNLNDLTSLFLLNGTDLTFHMIESAGVLQT 2023  
 QY 217 -GGS-----PPRSGTTLIRLVIDINDNVEPESEKTKVOPENS 255  
 Db 2024 RGSFTPREODDHEVAVEVRDNRVQKVAQALYRVSVDVNDNTEFQHLPTTYIQT 2083  
 QY 256 PVGSLVYVSANDLDTGSGNGELIVAF--YATERLTFTFRINSTGNLHLKALNEAIO 313  
 Db 2084 EPEDVLFQVSARDKIDGANGSVYGFADYA-----YFRIDPYVGDISTLKKPFYQALN 2137  
 QY 314 TYTLTQAKDGGGSGKCTVYVHTDINDNPELLMSLTSPIDENSPETVAVAFRIKDR 373  
 Db 2138 KHLKRIANDSGIPPLQTEVEVHTVRKNSNPLFQSPYKVKVPNT--TLTPIIHTQA 2195  
 QY 374 DSGNNAKMCISIODHLPEVLPKSPVENEFTYLTALRDEERTEYNITITVD--LGTPL 431  
 Db 2196 RSPGRLIYINVEEPLMLFTTDRKTVLYTGGLDSEKKNHFTYRAIDTALGS--- 2252  
 QY 432 KQHNLITVYSDVNDNAPTFSGTTYTLKVRNNSPALHIGVSATDRDSCGANAQYTSLL 491  
 Db 2253 FSEAVVELVEDINDNPPFESOLVYTSSESPQATFVIOILLASDDSGONOVSYOIV 2312  
 QY 492 PPHDQPLPGSLVS-----INADGOLFALRSIDFALAFERFGAARSPALSSQALY 547  
 Db 2313 ED-----GDSVSKFFRINSGTGEITFQIELDYTHGHFKYKRAADKGPPLTGETLV 2365  
 QY 548 RVLVADANDNAPFVLYPLONGSAPCTELVPRAEAGYLVAVVADGSGONAMLSYOLLA 607  
 Db 2366 VVNVSDINDNPPKPREPOYAN-----VSELATCGHLVLAQVOLDPDIOTSLREYLIL 2419  
 QY 608 KATEPGLFGVMAHNG 622  
 Db 2420 SGNDRHFSINSTSG 2434

## RESULT 11

T15276  
 hypothetical protein R10P2.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T15276  
 R:Du, Z.; Gattung, S.  
 submitted to the EMBL data library, May 1997  
 A:Description: The sequence of C. elegans cosmid R10P2.  
 A:Reference number: Z18320  
 A:Accession: T15276  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2163 <DUZ>  
 A:Cross-references: EMBL:AF003388; NID:92088850; PID:92088852; PIDN:AB54266.1; GSPDB  
 A:Experimental source: strain Bristol NZ; clone R10P2  
 C:Genetic source: C. elegans  
 A:Gene: CESP.R10P2.1  
 A:Map position: 3  
 A:Introns: 34/1; 114/3; 735/3; 1552/2; 1800/3; 1865/3; 1911/3; 2026/3; 2108/2; 2150/3

## Query Match

Best Local Similarity 15.1%; Score 485; DB 2; Length 2163;  
 Pred. No. 1.4e-23;  
 Matches 149; Conservative 82; Mismatches 224; Indels 60; Gaps 16;

QY 123 VRDINDSPVFLDREITNLITESTPGATFLLESASHDSVGINNLNRTISSNVFHN 181  
 Db 46 LKDVNDMPFLFYPNEHTLYRBCGPKPEEPLLVASDMSDGTGEVSYHLLSESSFSFIN 105



```

F:28-166/Domain: propeptide #status predicted <PRO>
F:167-913/Product: R-cadherin #status predicted <MAT>

```

F:167-721/Domain: extracellular #status predicted <EXT>  
 F:169-274/Domain: cadherin repeat homology <CR1>  
 F:244-249/Region: cadherin binding #status predicted  
 F:277-389/Domain: cadherin repeat homology <CR2>  
 F:300-304/Domain: calcium binding #status predicted <CAB>  
 F:392-504/Domain: cadherin repeat homology <CR3>  
 F:507-612/Domain: cadherin repeat homology <CR4>  
 F:613-721/Domain: cadherin repeat homology <CR5>  
 F:722-753/Domain: transmembrane #status predicted <TM>  
 F:754-913/Domain: intracellular #status predicted <INT>  
 F:870-885/Region: serine-rich  
 F:283,412,557,632,661,702/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.9% Score 445.5; DB 2; Length 916;  
 Best Local Similarity 26.2%; Pred. No. 1.6e-21;

Matches 157; Conservative 81; Mismatches 210; Indels 151; Gaps 21;

QY 129 HSPVLEDEITLLESTPGATFLESADSDVGI-NLNNTYISSNVYEHINVDNGE 187  
 DB 131 HKPQKGVVLD--PSPPKDTLLPMPQHONANGLRRKKDWYTP-----INVPENSR 183  
 QY 168 GNVYSELVDKYLDEEPELRLITGLDGSSP-----RSGTTLIR----- 229  
 DB 194 GPFPQOLVRISDKNDIP-IRYSITGVADQPMVEVFSINSMGRMYVTRPMDREEHAS 242  
 QY 230 -----ILYLDINDNVPFEVSELYVQVPEVSPVGLVVTYSAR 267  
 DB 243 YHLRAHVDNMGNKVENPIDLYIVDMNDNHPFINOVINCSDSKPTVMTTAN 302  
 QY 268 DLD--TGSNGEIVYAFYATER--TLKTRINSTGNL-HLKAELENIQTOTYLTIOAK 322  
 DB 303 DADSTTANGVRYRIYQTQSPQNMFTINSEIGDIYVAAQMDREKVOQYTVIOAT 362  
 QY 323 D-----GGSLGKCTVYVHTVDINDNPELLMSLTSPENSEYVAVFRIDRDSGN 377  
 DB 363 DMENGLNGLSTNTAIIITVDVNDNPEFTASTAGEVENSVEYVAVMLTVVDRD--- 419  
 QY 378 NAKWVCSIODHLP-----FLKPSVENFYTLTERALDREERTENIT 420  
 DB 420 -----QHPSPMMNAVYRIISGDPGHSVRTPDTNMGMYVKAAYDELNRAPMLT 471  
 QY 421 ITYVDLG-----TPRLKTOHNLTVYSDVNDNAPFSQTYTLRVRENSPALHIGSVS 474  
 DB 472 VMVSNQAPLASIGMSFSQSTAGVTISIMDINE-APYFSPNHKLIRIEEGVPGLVITFS 530  
 QY 475 AFDDSGANAOVTYSLPPHPDQLPLGLSVINADNGOLFALSLDPEAL-----QAFER 530  
 DB 531 AVDPDRMQQAVRYSKL--SDP-----ASMLHINATNGOITTVAVLDRESLYTKNNYEAT 584  
 QY 531 VGAADRGSPALISQALVRLVADANDNAPFVLYPLONGSAPCTELVRAAE-----AGYLV 586  
 DB 585 FLADNGIIPRPSGCTGLQIYLIDINDNAP-----ELLPKEAQIDERRPLNA 630  
 QY 587 AKVAVDGD-----SGQAMLSYOLKATEPGLFVG 617  
 DB 631 INTAADADVHPNIGPYFELFPVPAVRKNMTITRLNGDYAOLSLIKIL-YLEAGMYDV 688

## RESULT 14

S44887

ZK112.7 protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 14-Sep-1994 #sequence\_revision 12-May-1995 #text\_change 23-Mar-2001

C:Accession: S44887

R:Du, Z.

submitted to the EMBL Data Library, May 1993

A:Description: Sequence of the C. elegans cosmid ZK112.

A:Reference number: S44816

A:Accession: S44887

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-3343 &lt;DU&gt;

A:Cross-references: EMBL:L14324; NID:g289740; PID:g289742

C:Genetics:  
 A:Introns: 39/1; 74/3; 122/3; 209/2; 287/3; 333/2; 489/1; 611/2; 714/1; 939/3; 1002/2  
 C:Keywords: cytoskeleton; transmembrane protein

Query Match 13.8% Score 442.5; DB 2; Length 3343;  
 Best Local Similarity 26.5%; Pred. No. 1.8e-20;

Matches 175; Conservative 94; Mismatches 249; Indels 143; Gaps 28;

QY 35 VAETERTGPLANLAIDLGLGVEELNARGCRISYDSEITGILLNPLTGLDNLNEXDRE 94  
 DB 1564 ISEADFTFVGKIVL-LGAGLEGSVFR-----IQDDNYNF-TISPDDGILFNPSLDFEN 1617  
 QY 95 LCGTEPCVLPFOLLLEKPEQIFRAELWVDINDHSPVELDRE-ITLNLIES-TPPGATE 152  
 DB 1618 -----IKTYRPNITAGKSTQVILIHVDEDEAPRFTTGVVNLKYLEELDVPYPL 1669  
 QY 153 LLES--AHSDVGINNLNNTYSS--NVFHNVDHNGGNGVYSELVDKYLDEEVPPL 208  
 DB 1670 ITGSSIAEDDEQNLVYTSILSGNTSLFVAVN--STIGDILSL---PLDRESSLA 1722  
 QY 209 RLTLTGIDGSPSPSGTTLIRILVLDINDNPEFEVSELYKQVPEVSPVGLVVTYSAR 268  
 DB 1723 ELIIEAKDAGIPSLASTSKILHVGIDINDNTEPEFELISFYIKISENSKIGSKILILATD 1782  
 QY 269 LDTGSGEIVYAFYATERPLKTRINSTGNLHLKAELENIQTOTYLTIOAKDGGLS 328  
 DB 1783 KD--KDAELOYSLESNDEITLI--PFRINAVATGWTYAGKVRRENEEFREFVAVTD-GEKS 1838  
 QY 329 GKCTVYVHTVDINDNPEEL--LMSSLTSPENSPETVAVFRIDR-----SGNNA 379  
 DB 1839 SKYIVIEIHVEDNDNHPMINDRNSDLEVPDPTRSYE-IIHVIVHDLKSDHLKFLSNS 1897  
 QY 380 KMWCS-----IOD-----HLPF----- 391  
 DB 1898 NNLNLSNGEITLKSPLQTAVPVRYVSDDAGVAFMEYLFHPSRKHHPVFEKIDTVSY 1957  
 QY 392 -----VAKPSYENF-YTLY-----TERALDREERTENI-TTV 423  
 DB 1958 REHDEQELAVFRANGDSIRYSIVSRCSDLHLEKSTGLIKTKSSIDAEYSCELVFIAT 2017  
 QY 424 TDLGTPRLKTOHNLTVYSDVNDNAPFSQTYTLRVRENSPALHIGSVATDSCAN 483  
 DB 2018 TYEDKNPSTIIRKATIKIYDINDNPRPDQILRRYVENSQPKL-IGHVIA--RDIDRS 2074  
 QY 484 AOYVTSIL--PPHPDQLPLGLSVINADNGOLFALSLDPEALQAFERVGAADRGSPA 540  
 DB 2075 SRVFEIYVGDANDNHEFV-----TESQIEBSYRDLDRKSEYHLIYEALIDGKPR 2125  
 QY 541 LSSQALVRLVADANDNAPFV--LYPLONGSAPCTELVRAAEGLYVAKVYVADGDSGQ 598  
 DB 2126 RRGNTTVYTVLDEDDNAPRFRIRFVE-----VPEDVRIGEPVLIQSLASDADERS 2176  
 QY 599 N 599  
 DB 2177 N 2177

## RESULT 15

G02678

cadherin-14 - human

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 21-Jan-2000

C:Accession: G02678

R:Shibata, T.; Shimoyama, Y.; Gotoh, M.; Hirohashi, S.

submitted to the EMBL Data Library, May 1996

A:Reference number: H01584

A:Accession: G02678

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-790 &lt;SHI&gt;

A:Cross-references: EMBL:U59325; NID:g1389852; PIDN:AA02933.1; PID:g1389853

C:Superfamily: cadherin; cadherin repeat homology  
 F:163-268/Domain: cadherin repeat homology <CDH>







GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 10:46:02 ; Search time 37 seconds  
(without alignments)  
3491.663 Million cell updates/sec

Title: US-10-050-704-125

Perfect score: 3208

Sequence: 1 MEARVHALQKRVSLCVF.....KATEPGLFGWAHNGEVRTA 627

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.vertibrate:\*  
14: sp.unclassified:\*  
15: sp.rvirts:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2526	78.7	723	11	Q925L6
2	2526	78.7	786	11	Q91Y04
3	2256	70.3	799	11	Q91X29
4	2234	70.3	800	4	Q96T99
5	2230	70.1	844	11	Q925L1
6	2240	69.8	776	4	Q96SE9
7	2167	67.5	799	11	Q91Y00
8	2164	67.5	784	11	Q91X27
9	2159	67.3	794	11	Q91X28
10	2157	67.2	784	11	Q925W6
11	2157	67.2	794	11	Q925L0
12	2155	67.2	792	11	Q91Y02
13	2148	67.0	796	11	Q91Y05
14	2138	66.6	792	11	Q925L3
15	2104	65.6	799	11	Q91Y08
16	2098	65.4	799	11	Q925L4

17	2078	64.8	792	11	Q91X25
18	2052.5	64.0	734	4	Q96T90
19	2050.5	63.9	774	11	Q91Y48
20	2042.5	63.7	798	11	Q925M2
21	2040.5	63.6	793	11	Q91X21
22	2036.5	63.5	798	11	Q91X23
23	2023.5	63.1	772	11	Q91X24
24	2023.5	63.1	796	11	Q91Y06
25	2013.5	62.8	797	11	Q91X22
26	2011.5	62.7	797	11	Q91Y28
27	2009.5	62.6	789	11	Q91Y07
28	2007	62.6	801	11	Q925L2
29	2007	62.6	801	11	Q91Y01
30	1996.5	62.2	784	11	Q91X26
31	1986.5	61.9	797	11	Q91Y05
32	1977.5	61.6	802	11	Q91Y03
33	1969.5	61.4	784	11	Q925M5
34	1879	58.6	744	4	Q9HAB4
35	1836	57.2	587	4	Q96T98
36	1723.5	53.7	818	11	Q91Y08
37	1693	52.8	593	11	Q925M0
38	1573	49.0	595	11	Q925M7
39	1432.5	44.7	935	11	Q91XY1
40	1428.5	44.5	931	11	Q91XY6
41	1423	44.4	932	11	Q91XY9
42	1418	44.2	813	4	Q9Y5D2
43	1418	44.2	931	4	Q9Y5G8
44	1415.5	44.1	850	4	Q9Y5E0
45	1415.5	44.1	936	4	Q9Y5H3

## ALIGNMENTS

## RESULT 1

ID Q925L6 PRELIMINARY: PRT: 723 AA.  
AC Q925L6;  
DT 01-DEC-2001 (TREMBLER, 19, Created)  
DT 01-DEC-2001 (TREMBLER, 19, Last sequence update)  
DT 01-JUN-2002 (TREMBLER, 21, Last annotation update)  
DE Protocadherin-beta (Fragment).  
GN PCDBH15 OR PCDBH7.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21223055; Pubmed-11322959;  
RA Vahdai K., Koolis P., Vanden Eynde E., van Roy F.;  
RT "The human and murine protocadherin-beta one-exon gene families show  
RT high evolutionary conservation, despite the difference in gene  
RT number.";  
RL FEBS Lett. 495:120-125(2001).  
CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
DR EMBL: AF326308; AAK53233.1;  
DR MGD: MGI:2136750; Pcdh15.  
DR InterPro: IPR002126; Cadherin.  
DR Pfam: PF00028; cadherin.5.  
DR PROSITE: PS00232; CADHERIN\_1; UNKNOWN\_5.  
DR PROSITE: PS0268; CADHERIN\_2; 6.  
KW Calcium-binding; Cell adhesion; Glycoprotein.  
FT NON\_TER 723  
FT SEQUENCE 723 AA; 79555 MW; 678543AB3647DEB6 CRC64;

Query Match 78.7%; Score 2526; DB 11; Length 723;  
Best Local Similarity 78.1%; Pred. No. 8.5e-167;  
Matches 489; Conservative 57; Mismatches 80; Indels 0; Gaps 0;

OY 1 MEARVHALQKRVSLCVFGLVSWAGAEPLRYVAETERTGFLANLADIGLGEELS 60  
DB 1 MEARVHALQKRVSLCVFGLVSWAGAEPLRYVAETERTGFLANLADIGLGEELS 60

```

QY 61 ARGCRTVSDETIGFLINPFLDGLLNKREELCGTEPCVLPFOILLERKPOIFRAE 120
DB 61 AEARIVSDQNTFRLLSPFLSLTGLNELINKREELCGTEPCVLPFOILLERKPOIFRAE 120
QY 121 LWRVINDHSPVFLDREITLNTLESTPGATFLLESADSDVGINLNRYTSSVNFHI 180
DB 121 LWRVINDHSPVFLDREITLNTLESTPGATFLLESADSDVGINLNRYTSSVNFHI 180
QY 181 NVHDNGEGNVSEVLVDKREVEPELRLTLTGIDGSGPPRSCTTLIRIIVLINDVNP 240
DB 181 NVHDNGEGNVSEVLVDKREVEPELRLTLTGIDGSGPPRSCTTLIRIIVLINDVNP 240
QY 241 EVELSELYKQVPEPNSPVSLVTVASARLDGSGEIVYAFYATERTLKTFRINSTGN 300
DB 241 EVELSELYKQVPEPNSPVSLVTVASARLDGSGEIVYAFYATERTLKTFRINSTGN 300
QY 301 LHLKAEINVEALQIYTLTIQAKDGGGLSGKCVVYVHTVDINDNPPELLMSSLTSPIPENS 360
DB 301 LHLKAEINVEALQIYTLTIQAKDGGGLSGKCVVYVHTVDINDNPPELLMSSLTSPIPENS 360
QY 361 PETVAVAFRIKRDGSGNNAKWVCSTIDHLPFLKPSVENFTYTERALDREERTENIT 420
DB 361 PETVAVAFRIKRDGSGNNAKWVCSTIDHLPFLKPSVENFTYTERALDREERTENIT 420
QY 421 ITVTDLGPRLKTOHNLTVYSDVNDNAPTFSQTTITLRENNSPALHIGSVATDRDS 480
DB 421 ITVTDLGPRLKTOHNLTVYSDVNDNAPTFSQTTITLRENNSPALHIGSVATDRDS 480
QY 481 GANAQVTVSLPRLPDPOLPLGSLVSINADNGQLFALSLDEALQAEFFRGADRGSPA 540
DB 481 GANAQVTVSLPRLPDPOLPLGSLVSINADNGQLFALSLDEALQAEFFRGADRGSPA 540
QY 541 LSSQALVAVLVADANDNAPEVLYPLONGSAPCTELVPRAAGYLVAKVAVDGDGSGNA 600
DB 541 LSSQALVAVLVADANDNAPEVLYPLONGSAPCTELVPRAAGYLVAKVAVDGDGSGNA 600
QY 601 WLSYQLKATEPGLFSGVVAHNGEVRT 626
DB 601 WLSYQLKATEPGLFSGVVAHNGEVRT 626

```

RESULT 2  
Q91X04 PRELIMINARY; PRT; 786 AA.

AC Q91X04;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Protocadherin beta 15.  
GN PCDHB15.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX MEDLINE=99308636; PubMed=10380929;  
RA Wu Q., Maniatis T.;  
RT "A striking organization of a large family of human neural cadherin-  
like cell adhesion genes.";  
RL Cell 97:779-790(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX MEDLINE=20202599; PubMed=10716726;  
RA Wu Q., Maniatis T.;  
RT "Large exons encoding multiple ectodomains are a characteristic  
feature of protocadherin genes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).  
RN [3]  
RP SEQUENCE FROM N.A.

```

RC TISSUE-BRAIN;  
RA MEDLINE=21154914; PubMed=11230163;  
RX Wu Q., Zhang T., Cheng J.-F., Kim Y., Grimwood J., Schmutz J.,  
RA Dickson M., Noonan J.P., Zhang M.Q., Myers R.M., Maniatis T.;  
RT "Comparative data sequence analysis of mouse and human protocadherin  
gene clusters.";  
RL Genome Res. 11:389-404(2001).  
CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
DR EMBL: AY013776; AKK26065.1; -;  
DR MGI: 2136750; Pcdhb15.  
DR InterPro: IPR002126; Cadherin.  
DR Pfam: PF00028; cadherin; 5.  
DR PROSITE: PS00232; CADHERIN_1; UNKNOWN_5.  
DR PROSITE: PS0268; CADHERIN_2; 6.  
KW Calcium-binding; Cell adhesion; Glycoprotein.  
SQ SEQUENCE 786 AA; 86297 MW; 0EA6C04986F6FA90 CRC64;  
  
Query Match 78.7%; Score 2326; DB 11; Length 786;  
Best Local Similarity 78.1%; Pred. No. 9, 6e-167;  
Matches 489; Conservative 57; Mismatches 80; Indels 0; Gaps 0;  
  
QY 1 MEARVYHALOKROYSLCYFLGVSAGAPLRYFAEETERGTFLANLAIDGLVEELS 60  
DB 1 MEARLEQAVQKROVLFVFLGASMASABOLQYSAEETERGTFLANLAKDGLGELS 60  
QY 61 ARGCRTVSDETIGFLINPFLDGLLNKREELCGTEPCVLPFOILLERKPOIFRAE 120  
DB 61 AEARIVSDQNTFRLLSPFLSLTGLNELINKREELCGTEPCVLPFOILLERKPOIFRAE 120  
QY 121 LWRVINDHSPVFLDREITLNTLESTPGATFLLESADSDVGINLNRYTSSVNFHI 180  
DB 121 LWRVINDHSPVFLDREITLNTLESTPGATFLLESADSDVGINLNRYTSSVNFHI 180  
QY 181 NVHDNGEGNVSEVLVDKREVEPELRLTLTGIDGSGPPRSCTTLIRIIVLINDVNP 240  
DB 181 NVHDNGEGNVSEVLVDKREVEPELRLTLTGIDGSGPPRSCTTLIRIIVLINDVNP 240  
QY 241 EVELSELYKQVPEPNSPVSLVTVASARLDGSGEIVYAFYATERTLKTFRINSTGN 300  
DB 241 EVELSELYKQVPEPNSPVSLVTVASARLDGSGEIVYAFYATERTLKTFRINSTGN 300  
QY 301 LHLKAEINVEALQIYTLTIQAKDGGGLSGKCVVYVHTVDINDNPPELLMSSLTSPIPENS 360  
DB 301 LHLKAEINVEALQIYTLTIQAKDGGGLSGKCVVYVHTVDINDNPPELLMSSLTSPIPENS 360  
QY 361 PETVAVAFRIKRDGSGNNAKWVCSTIDHLPFLKPSVENFTYTERALDREERTENIT 420  
DB 361 PETVAVAFRIKRDGSGNNAKWVCSTIDHLPFLKPSVENFTYTERALDREERTENIT 420  
QY 421 ITVTDLGPRLKTOHNLTVYSDVNDNAPTFSQTTITLRENNSPALHIGSVATDRDS 480  
DB 421 ITVTDLGPRLKTOHNLTVYSDVNDNAPTFSQTTITLRENNSPALHIGSVATDRDS 480  
QY 481 GANAQVTVSLPRLPDPOLPLGSLVSINADNGQLFALSLDEALQAEFFRGADRGSPA 540  
DB 481 GANAQVTVSLPRLPDPOLPLGSLVSINADNGQLFALSLDEALQAEFFRGADRGSPA 540  
QY 541 LSSQALVAVLVADANDNAPEVLYPLONGSAPCTELVPRAAGYLVAKVAVDGDGSGNA 600  
DB 541 LSSQALVAVLVADANDNAPEVLYPLONGSAPCTELVPRAAGYLVAKVAVDGDGSGNA 600  
QY 601 WLSYQLKATEPGLFSGVVAHNGEVRT 626  
DB 601 WLSYQLKATEPGLFSGVVAHNGEVRT 626

```

RESULT 3  
Q91X29 PRELIMINARY; PRT; 799 AA.

AC Q91X29;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Protocadherin beta 20.  
 GN PCDB20.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE=99308636; PubMed=10380929;  
 RA Wu Q., Maniatis T.;  
 RT "A striking organization of a large family of human neural cadherin-  
 RT like cell adhesion genes";  
 RL Cell 97:779-790(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE=20202599; PubMed=10716726;  
 RA Wu Q., Maniatis T.;  
 RT "Large exons encoding multiple ectodomains are a characteristic  
 RT feature of protocadherin genes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE=21154914; PubMed=11230163;  
 RA Wu Q., Zhang T., Cheng J.-F., Kim Y., Grimwood J., Schmutz J.,  
 RA Dickson M., Noonan J.P., Zhang M.Q., Myers R.M., Maniatis T.;  
 RT "Comparative dna sequence analysis of mouse and human protocadherin  
 RT gene clusters";  
 RL Genome Res. 11:389-404(2001).  
 CC -1- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.  
 DR EMBL: AY013782; AAK26071.1;  
 DR MGD: MGI:2136758; Pcdh20.  
 DR InterPro: IPR002106; AATRNA\_1ligase1.  
 DR Pfam: PF00028; cadherin. 5.  
 DR PROSITE: PS00179; AA\_TRNA\_LIGASE\_II.1; UNKNOWN\_1.  
 DR PROSITE: PS00232; CADHERIN\_1; UNKNOWN\_5.  
 DR PROSITE: PS50268; CADHERIN\_2; 5.  
 DR Calcium-binding; Cell adhesion; Glycoprotein.  
 SQ SEQUENCE 799 AA; 87800 MW; 99C7A595F1673147 CRC64;  
 Query Match 70.3%; Score 2256; DB 11; Length 799;  
 Best Local Similarity 68.9%; Pred. No. 5.3e-148;  
 Matches 426; Conservative 91; Mismatches 101; Indels 0; Gaps 0;  
 QY 9 LOKROYSLICVFLGYSMAGAEPLRYFAETRGTFIANLAIDLGLVGEELSARGCRIYS 68  
 DB 9 LOKROYLVEFVFLGLSQASAEISRYSAVABETELGSFVANLAKDLGLVGEELSREARYVS 68  
 QY 69 DETIGLLNPLTGDLLEKLDREELCGTEPCVLPQOLLLEKPPQIFRAELWVDIND 128  
 DB 69 DDNRKHLNLNLGTMDLNERLDRDLRGSTQPCVLPFOVVENPLQFRAELHVRDIND 128  
 QY 129 HSPVFDREITLILSTPGATFLESADSDVGINNRTTISSNVFEHINVDNEGN 188  
 DB 129 HSTFTLDEKLETTIKISSTTIGATFLENQDDIGSNSLODDISISSFFYKIHDSGCG 188  
 QY 189 NVYSELVLDKVDREVEPELRITLTLGIDGSPPRSCTTLIRILVLDINDNVEFEVSLYK 248  
 DB 189 KIYPELVLDKALDRHEESSELRLTLTALDGGSPPRSCTTLIVIKVLDINDNAPFAQSYVE 248  
 QY 249 VQVPEPSPGSLVYVSARDLDTGSGEIVAFYATERTKTRFINSNGNLKAEIN 308  
 DB 249 VQVPEPSPGSLVYVSARDLDTGSGEIVAFYATERTKTRFINSNGNLKAEIN 308  
 QY 309 YEAIOYTLTLIAKDGSGSGCKTVVHTDINDNPELIMSLSLTPFENSPFIYVAV 368  
 DB 309 FEYIOGTVNIOATDGGGSAKCTISVKYVLDINDNAPVIMSVTATPEMASETIVALE 368  
 QY 369 RIRDRSGNNAKVCSTODHLPFVLRPSVENFTYVTERALDRERTEVNITITVDTLGT 428  
 DB 369 RIRDRSGNNAKVCSTODHLPFVLRPSVENFTYVTERALDRERTEVNITITVDTLGT 428

DB 369 SVRDSDGSDNGRILCSIODHLPLIKPSKFNFTLLSEKALDRERREAVNITITASDMGT 428  
 QY 429 PRKTOHNLVYVSVDVNDNAPEFSQTYTLRVENNSPALHIGSVATDRSGANDQVY 488  
 DB 429 PRKTOHNLVYVSVDVNDNAPEFSQTYTLRVENNSPALHIGSVATDRSGANDQVY 488  
 QY 489 SLPPHPDPLPLGSLVSIANDNGOLFALRSIDFALQAFERYGADRGSPALSSQALVR 548  
 DB 489 SLPPHPDPLPLGSLVSIANDNGOLFALRSIDFALQAFERYGADRGSPALSSQALVR 548  
 QY 549 VLVADANDNAPEVLYPLONGSAPCTELVPPRAEAGYVAKVAVAVDQSGQANVLSTQLK 608  
 DB 549 VLVADANDNAPEVLYPLONGSAPCTELVPPRAEAGYVAKVAVAVDQSGQANVLSTQLK 608  
 QY 609 ATEPGLFVNAHNGEVRT 626  
 DB 609 ATEPGLFVNAHNGEVRT 626  
 RESULT 4  
 ID Q96T99 PRELIMINARY; PRT; 800 AA.  
 AC Q96T99;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Protocadherin-beta10.  
 GN PCDB10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21223035; PubMed=11322959;  
 RA Vanhaest K., Kools P., Vanden Eynde E., van Roy F.;  
 RT "The human and murine protocadherin-beta one-exon gene families show  
 RT high evolutionary conservation, despite the difference in gene  
 RT number";  
 RL FEBS Lett. 495:120-125(2001).  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 DR EMBL: AF217748; AAK51616.1;  
 DR InterPro: IPR002106; AATRNA\_1ligase1.  
 DR Pfam: PF00028; cadherin. 5.  
 DR PROSITE: PS00179; AA\_TRNA\_LIGASE\_II.1; UNKNOWN\_1.  
 DR PROSITE: PS00232; CADHERIN\_1; UNKNOWN\_5.  
 DR PROSITE: PS50268; CADHERIN\_2; 6.  
 DR Calcium-binding; Cell adhesion; Glycoprotein.  
 SQ SEQUENCE 800 AA; 87537 MW; 7DF4B52E0B78A263 CRC64;  
 Query Match 70.3%; Score 2254; DB 4; Length 800;  
 Best Local Similarity 70.6%; Pred. No. 7.4e-148;  
 Matches 436; Conservative 70; Mismatches 112; Indels 0; Gaps 0;  
 QY 10 OKROYSLICVFLGYSMAGAEPLRYFAETRGTFIANLAIDLGLVGEELSARGCRIYS 69  
 DB 10 OKROYLVEFVFLGLSQASAEISRYSAVABETELGSFVANLAKDLGLVGEELSREARYVS 69  
 QY 10 ROKROYLVEFVFLGLSQASAEISRYSAVABETELGSFVANLAKDLGLVGEELSREARYVS 69  
 DB 10 ROKROYLVEFVFLGLSQASAEISRYSAVABETELGSFVANLAKDLGLVGEELSREARYVS 69  
 QY 70 ETIGLLNPLTGDLLEKLDREELCGTEPCVLPQOLLLEKPPQIFRAELWVDIND 129  
 DB 70 DDNRKHLNLNLGTMDLNERLDRDLRGSTQPCVLPFOVVENPLQFRAELHVRDIND 129  
 QY 130 SPVFDREITLILSTPGATFLESADSDVGINNRTTISSNVFEHINVDNEGN 189  
 DB 130 SPVFDREITLILSTPGATFLESADSDVGINNRTTISSNVFEHINVDNEGN 189  
 QY 130 AVEFQDKETVLIKISNTTIGATFLENQDDIGSNSLODDISISSFFYKIHDSGCG 189  
 DB 130 AVEFQDKETVLIKISNTTIGATFLENQDDIGSNSLODDISISSFFYKIHDSGCG 189  
 QY 190 VYSELVLDKVDREVEPELRITLTLGIDGSPPRSCTTLIRILVLDINDNVEFEVSLYK 249  
 DB 190 VYSELVLDKVDREVEPELRITLTLGIDGSPPRSCTTLIRILVLDINDNVEFEVSLYK 249  
 QY 249 IYPELVLDKALDRHEESSELRLTLTALDGGSPPRSCTTLIVIKVLDINDNAPFAQSYVE 309  
 DB 249 IYPELVLDKALDRHEESSELRLTLTALDGGSPPRSCTTLIVIKVLDINDNAPFAQSYVE 309  
 QY 250 QVPEPSPGSLVYVSARDLDTGSGEIVAFYATERTKTRFINSNGNLKAEIN 309  
 DB 250 QVPEPSPGSLVYVSARDLDTGSGEIVAFYATERTKTRFINSNGNLKAEIN 309

Db 250 QAPENSIGFLIVKWAEDVDGNAEVSYSFDSASENRTTFQINPESGELFELLDY 309  
 QY 310 EAIQYITLTOAKDGGSLGKCTVYVHTDINDNPELMSLSPIPENSPEVAVR 369  
 Db 310 ELVNSTKINIQAMDGGGSAKRCVLEVDYDNDNPELIVSSFSNSAENSPEPIAVK 369  
 QY 370 IHDRDSGNNAKVCSTIOHLPVKPSVENFTLVTERALDREETEVNITITVDTCTP 429  
 Db 370 INDRSGEKNKVCYIOENLPFLKPSVENFTLVTERALDREETEVNITITVDTCTP 429  
 QY 430 RLKTHNLTVTVSDVNDNAPFTFSQTTTLRVRENNSPALHIGSVSATDRSGANAQVTS 489  
 Db 430 RLKTHNLTVTVSDVNDNAPFTFSQTTTLRVRENNSPALHIGSVSATDRSGANAQVTS 489  
 QY 490 ILPPDLPOLPLGSLVSTINADNGOLFALSLDFEALQAEFEFVGADRGSPALSSQALVR 549  
 Db 490 ILPPDLPOLPLGSLVSTINADNGOLFALSLDFEALQAEFEFVGADRGSPALSSQALVR 549  
 QY 550 LVADANDNAPFLVYLPONGSAPCTELVPRAAEAGYLVKVAVDGSGONAMLSYOLKA 609  
 Db 550 LVADANDNAPFLVYLPONGSAPCTELVPRAAEAGYLVKVAVDGSGONAMLSYOLKA 609  
 QY 610 TEPGLFGVWANGEVRTA 627  
 Db 610 TEPGLFGVWANGEVRTA 627

## RESULT 5

Q925L1 PRELIMINARY; PRT; 844 AA.  
 AC Q925L1;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Protocadherin-beta.  
 CN PCDB20 OR PCDB14.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RA MEDLINE=2123055; PubMed=11322959;  
 RA Vanhaest K., Kools P., Vanden Eynde E., van Roy F.;  
 RT "The human and murine protocadherin-beta one-exon gene families show  
 RT high evolutionary conservation, despite the difference in gene  
 RT number."  
 RL FBS Lett. 495:120-125(2001).  
 CC -1- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.  
 DR EMBL: AF236313; AAK53238.1;  
 DR MGD: MGI:216758; Pcdhb20.  
 DR InterPro: IPR002106; AATRNA\_LigaseII.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin; 5.  
 DR PROSITE: PS00179; AA\_TRNA\_Ligase\_II\_1; UNKNOWN\_1.  
 DR PROSITE: PS00232; CADHERIN\_1; UNKNOWN\_5.  
 DR PROSITE: PS00268; CADHERIN\_2; 5.  
 KM Calcium-binding; Cell adhesion; Glycoprotein.  
 SQ SEQUENCE 844 AA; 93034 MW; 303E780FF2E9E7C CRC64;

Query Match 70.1%; Score 2250; DB 11; Length 844;  
 Best local similarity 68.0%; Pred. No. 1.5e-147;  
 Matches 425; Conservative 94; Mismatches 106; Indels 0; Gaps 0;

QY 2 EAVVAHALOKROYSLICVFLGSMWAGAEPLRIVAEETRGFLANLAIDLGLVEELSA 61  
 Db 47 EGGGLSSCRKROYLVEFVGLGSAESLRYSAEETIGSFVANLARDLGVAELSS 106  
 QY 62 RGRIVSDERTIGFLINPLTGLDNLNEKIDREELGPTPCVLPFOLLLEKFPQJFRAEL 121  
 Db 107 REARVYSDDKKHLNLTGLDNLNEKIDREELGPTPCVLPFQVYVLENLQFVRAEL 166

QY 122 WVRDINDHSPVFLDREITLILESTPGATFLESADSDVGINLNNYNTISSNVFHN 181  
 Db 167 HVRIIDHSPFLDKETIKISTTIGATFLEENQDDIGNSLQDVISISSFFYK 226  
 QY 182 VHDGEGNNVSELYLKVLDREVPELRLTLTGDDGSPRSGTTLIRLVLDINDNVE 241  
 Db 227 IHSDGDKITPELVLDALDHEESESELRLTLTALDGSPPRSGTTLIRLVLDINDNVE 286  
 QY 242 FVESLYVQVPENSPVGSLVTVSARDLDTGSGEIVYAFVTEERTLTFRINSNGM 301  
 Db 287 FASQFVQVPEMDPRIGSSITTAISADLDMGITGKISTYFLASEDIRKTFEINPTSGEV 346  
 QY 302 HLAELNVEALIQYITLTOAKDGGSLGKCTVYVHTDINDNPELMSLSPIPENS 361  
 Db 347 NLRSLMFEVYIQSYVNIQATDGGGSAKCTLSVKVLINDNAPETVMSVKAIDENS 406  
 QY 362 ETVYVAFRITDRSGNNAKVCSTIOHLPVKPSVENFTLVTERALDREETEVNIT 421  
 Db 407 ETVYVAFRITDRSGNNAKVCSTIOHLPVKPSVENFTLVTERALDREETEVNIT 466  
 QY 422 TVTDLGTPRLKTOHNTLVYSDVNDNAPFTFSQTTTLRVRENNSPALHIGSVSATDRSG 481  
 Db 467 TASDKGTPLRTTOHTIRVOYSDINDNAPFTOTSTYTFYENNSPALHIGTSATDKDSG 526  
 QY 482 ANAQVYSLPPLPDPOLPLGSLVSTINADNGOLFALSLDFEALQAEFEFVGADRGSPAL 541  
 Db 527 SNAHITVSLPADDPELALASISTINADNGOLFALRALDYLQAEFEFVGATDRSGPAL 586  
 QY 542 SSOALVRVLYADANDNAPFLVYLPONGSAPCTELVPRAAEAGYLVKVAVDGSGONAM 601  
 Db 587 SSOALVRVLYADANDNAPFLVYLPONGSAPCTELVPRAAEAGYLVKVAVDGSGONAM 646  
 QY 602 LSYOLKATEPGLFGVWANGEVRT 626  
 Db 647 LSYOLKATEPGLFGVWANGEVRT 671

## RESULT 6

Q96SE9 PRELIMINARY; PRT; 776 AA.  
 AC Q96SE9;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Protocadherin beta 8a.  
 CN PCDB8A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99308636; PubMed=10380929;  
 RA Wu Q., Maniatis T.;  
 RT "A striking organization of a large family of human neural cadherin-  
 RT live cell adhesion genes."  
 RL Cell 97:779-790(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20202599; PubMed=10716726;  
 RA Wu Q., Maniatis T.;  
 RT "Large exons encoding multiple ectodomains are a characteristic  
 RT feature of protocadherin genes."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21154914; PubMed=11230163;  
 RA Wu Q., Zhang T., Cheng J.-F., Kim Y., Grimwood J., Schmutz J.,  
 RA Dickson M., Noonan J.P., Zhang M.Q., Myers R.M., Maniatis T.;  
 RT "Comparative DNA sequence analysis of mouse and human protocadherin  
 RT gene clusters."  
 RL Genome Res. 11:389-404(2001).  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.







DT	01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE	Protocadherin-betaV.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RX MEDLINE=21223055; PubMed=11322959;  
RA Vanhalst K., Koole P., Vanden Eynde E., van Roy F.;  
RT "The human and murine protocadherin beta one-exon gene families show  
RT high evolutionary conservation, despite the difference in gene  
RT number.";  
RL FEBS Lett. 495:120-125(2001).  
RC -1- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.  
RL EMBL: AF326315; AK53240.1; -.  
DR DR  
DR MGD; MGI:2136760; Pcdhb22.  
DR InterPro; IPR002126; Cadherin.  
DR Pfam; PF00028; cadherin. 5.  
DR PROSITE; PS00232; CADHERIN\_1; UNKNOWN\_4.  
DR PROSITE; PS50268; CADHERIN\_2; 5.  
KM Calcium-binding; Cell adhesion; Glycoprotein.  
SQ SEQUENCE 794 AA; 86907 MW; 8B369FBB339595B7 CRC64;



DR InterPro: IPR002126; Cadherin.  
 DR InterPro: IPR003580; Protachykinin.  
 DR Pfam: PF00028; cadherin.5.  
 DR SMART: SM00203; TK; 1.  
 DR PROSITE: PS00232; CADHERIN\_1; UNKNOWN\_5.  
 DR PROSITE: PS50268; CADHERIN\_2; 6.  
 DR Calcium-binding; Cell adhesion; Glycoprotein.  
 KW SEQUENCE 792 AA; 86897 MW; F7CA3262A89E1D66 CRC64;

Query Match 66.6%; Score 2138; DB 11; Length 792;  
 Best Local Similarity 66.6%; Pred. No. 8.1e-140;  
 Matches 417; Conservative 80; Mismatches 129; Indels 0; Gaps 0;

QY 1 MEARVVALQKQVSLCFLVGVNAGAPLRYFAEETERTGFLANLAIDLGLVEELS 60  
 DB 1 MAARSCSVSRKQVLFLLGLGLAGSELGRYSVETERTGSLVANLANGLGVALA 60  
 QY 61 ARGCTVDEITGFLPLPLTGDLINLREELCGTEPCVLPQLLLEKPPQITRAE 120  
 DB 61 AKRRTVVDCKKHLFLDSHGDLITNEKDLCKGTEPCMLYFQILMDNPQITRAE 120  
 QY 121 LVRNDINSHSPVFLDREITLNTLESTPGATFLLESADSDVGGINNRYNTISSNYEHI 180  
 DB 121 LRILIDINSHSPVFLDREITLNTLESTPGATFLLESADSDVGGINNRYNTISSNYEHI 180  
 QY 181 NVHNGEGNYSSELYLQKVDREVEPELRLTLGLDGGSPRSGTTLRLIYLDINDNP 240  
 DB 181 PNHNSDEMTPEYLDALDMEQPESTLTLTALDGGSPRSGTTLRLIYLDINDNP 240  
 QY 241 EFVESLYKVQVSPNSVSLVTVSARDLDTGSGEIVYAFVYETKTLKTRFNISGN 300  
 DB 241 QPQELVEIQAPNSPIGLVYIKYTGEDVDGVAELSYSEFDASEDIRATFOINPSSGE 300  
 QY 301 LHLKALNVEAIQVYTLTQAKDGGSGKCTVVAVHTDINDPELMSLSLTPSPENS 360  
 DB 301 IITKALIDYEIKSYKINQVAVDGGGLSARCTVLRVLVDNDAPELIMSLVNEVENS 360  
 QY 361 PETVAVAFRIRDRSGNNNAKWCISIOHLEPVLPKPSVENEYTLVTEALDREERTENIT 420  
 DB 361 PETVAVAFRIRDRSGNNNAKWCISIOHLEPVLPKPSVENEYTLVTEALDREERTENIT 420  
 QY 421 ITVTDLGPRLKQHNLTVTVDVNDNAPTFSTYTLRYRENNSPALHIGSAYDRDS 480  
 DB 421 ITVSDLGTPRLTQHTIRVQVSDINDNAPAFNOSTYTLRYRENNSPALHIGTISADSDA 480  
 QY 481 GANQVYSLIPHPDPLPGSLVSNADNGQLFALSLDQFALQAEFFVGAADRQSPA 540  
 DB 481 GSNHITYSLIPDQDMLTSLVSNADNGQLFALSLDQFALQAEFFVGAADRQSPA 540  
 QY 541 LSSQALVRYLVADANDAPVLPVPIONGSAPCEIYPRAEAGYLVAKYAVVDGSGQA 600  
 DB 541 LSSQALVRYLVADANDAPVLPVPIONGSAPCEIYPRAEAGYLVAKYAVVDGSGQA 600  
 QY 601 WLSTQLLKATEPGLFGVMAHNGEVRT 626  
 DB 601 WLSTQLLKATEPGLFGVMAHNGEVRT 626  
 RESULT 15  
 QY1VD8 PRELIMINARY; PRT; 799 AA.  
 AC QY1VD8;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Protocadherin beta 17.  
 GN PCDB17.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN;  
 RX MEDLINE-99308636; PubMed-10380929;  
 RA Wu Q., Maniatis T.;  
 RT "A striking organization of a large family of human neural cadherin-  
 RT like cell adhesion genes.";  
 RL Cell 97:779-790(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE-20202599; PubMed-10716726;  
 RA Wu Q., Maniatis T.;  
 RT "Large exons encoding multiple ectodomains are a characteristic  
 RT feature of protocadherin genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE-21154914; PubMed-11230163;  
 RA Wu Q., Zhang T., Cheng J.-F., Kim Y., Grimwood J., Schmutz J.,  
 RA Dickson M., Noonan J.P., Zhang M.O., Myers R.M., Maniatis T.;  
 RT "Comparative dna sequence analysis of mouse and human protocadherin  
 RT gene clusters.";  
 RL Genome Res. 11:369-404(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-EYE, AND RETINA;  
 RA Strausberg R.;  
 RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 DR EMBL: AY013778; AKR26067.1; -;  
 DR EMBL: BC017149; AAH17149.1; -;  
 DR MGD: MGI:2136754; Pcdh17.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin.5.  
 DR PROSITE: PS00232; CADHERIN\_1; UNKNOWN\_4.  
 DR PROSITE: PS50268; CADHERIN\_2; 6.  
 DR Calcium-binding; Cell adhesion; Glycoprotein.  
 KW SEQUENCE 799 AA; 87743 MW; 13807F0927972952 CRC64;

Query Match 65.6%; Score 2104; DB 11; Length 799;  
 Best Local Similarity 65.9%; Pred. No. 1.9e-137;  
 Matches 406; Conservative 93; Mismatches 127; Indels 0; Gaps 0;

QY 1 MEARVVALQKQVSLCFLVGVNAGAPLRYFAEETERTGFLANLAIDLGLVEELS 60  
 DB 1 METAWMCNLRQVAFVLLHVSAGAEIGPYSIEETERTGSPFANLGRGLVDLEIS 60  
 QY 61 ARGCTVDEITGFLPLPLTGDLINLREELCGTEPCVLPQLLLEKPPQITRAE 120  
 DB 61 NRRARIISQENKEHLQNLQSGDLINLREELCGTEPCVLPQLLLEKPPQITRAE 120  
 QY 121 LVRNDINSHSPVFLDREITLNTLESTPGATFLLESADSDVGGINNRYNTISSNYEHI 180  
 DB 121 LVRNDINSHSPVFLDREITLNTLESTPGATFLLESADSDVGGINNRYNTISSNYEHI 180  
 QY 181 NVHNGEGNYSSELYLQKVDREVEPELRLTLGLDGGSPRSGTTLRLIYLDINDNP 240  
 DB 181 VTRNRSDEKRYPELVELEKELDREPEPELRLTLGLDGGAPRSGTAOVLIEVDINDNP 240  
 QY 241 EFVESLYKVQVSPNSVSLVTVSARDLDTGSGEIVYAFVYETKTLKTRFNISGN 300  
 DB 241 KFOQPTVQVQIPENSPGSLVTVSANDLSDGDKVLYALSDQSEDSISTLEVNPTGE 300  
 QY 301 LHLKALNVEAIQVYTLTQAKDGGSGKCTVVAVHTDINDPELMSLSLTPSPENS 360  
 DB 301 IRLKEVDFERTIPSEYVDIKATDGGGSLGKCTLLKAYVDVNDNAPPEMLALSIPVENS 360  
 QY 361 PETVAVAFRIRDRSGNNNAKWCISIOHLEPVLPKPSVENEYTLVTEALDREERTENIT 420  
 DB 361 PDEVVAVAVSVPDPSANGKMAISIEEDLPFLKSSKNFTYLVTKRALDREERTENIT 420  
 QY 421 ITVTDLGPRLKQHNLTVTVDVNDNAPTFSTYTLRYRENNSPALHIGSAYDRDS 480  
 DB 421 ITVTDLGPRLKQHNLTVTVDVNDNAPTFSTYTLRYRENNSPALHIGSAYDRDS 480

```

Db 421 ITVSDLGTPRLTQHTITVQVSDTNDNAPAFNQTSTLFEVRENNSPAMHIGTISATDSDA 480
OY 481 CANAQVYTSLLPPIHPDPLGSLYSINADNGOLFALNSLDFEALQAFEFYRGADRGSPA 540
    | : : : | | | | | | : : : | | | | | | : : : | | | | | |
Db 481 GSNHSISYSLPSHDPOLADSLISINADNGOLFALRALDEALQAFEFHGAIDQSPA 540
OY 541 LSSQALVAVLVADANDNAPFVLVPLONGSAPCTELVPRAEAGYLVAKVAVADGDSGONA 600
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 541 LSSQALVAVLVADNDNAPFVLVPMQNSAPCTELPRAAEPGYLVTKVAVADRDGONA 600
OY 601 WLSYQOLKATEPGLFGVAHNGEVRT 626
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 601 WLSFOLKATEPGLFSVAHNGEVRT 626

```

Search completed: May 16, 2003, 10:48:22  
 Job time : 41 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 10:46:02 ; Search time 24 Seconds  
(Without alignments)  
2519.777 Million cell updates/sec

Title: US-10-050-704-125

Perfect score: 3208

Sequence: 1 MEARVHALQKROVSLICVF.....KATEPGLFGVMAHNGEVRTA 627

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 362588 seqs, 96450795 residues

Total number of hits satisfying chosen parameters: 362588

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA\*

1: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubppaa/PCR\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3208	100.0	627	US-10-050-704-125	Sequence 125, App
2	2249	70.1	800	US-10-036-041-52	Sequence 52, App1
3	2249	70.1	800	US-10-035-855-52	Sequence 478, App
4	2249	70.1	800	US-10-174-590-478	Sequence 478, App
5	2249	70.1	800	US-10-176-758-478	Sequence 478, App
6	2249	70.1	800	US-10-175-737-478	Sequence 478, App
7	2249	70.1	800	US-10-173-706-478	Sequence 478, App
8	2249	70.1	800	US-10-175-738-478	Sequence 478, App
9	2249	70.1	800	US-10-175-752-478	Sequence 478, App
10	2249	70.1	800	US-10-176-482-478	Sequence 478, App
11	2249	70.1	800	US-10-176-757-478	Sequence 478, App
12	2249	70.1	800	US-10-176-913-478	Sequence 478, App
13	2249	70.1	800	US-10-180-552-478	Sequence 478, App
14	2249	70.1	800	US-10-180-557-478	Sequence 478, App
15	2249	70.1	800	US-09-931-836-52	Sequence 52, App1
16	2249	70.1	800	US-10-173-700-478	Sequence 478, App
17	2249	70.1	800	US-10-174-572-478	Sequence 478, App
18	2249	70.1	800	US-10-174-579-478	Sequence 478, App
19	2249	70.1	800	US-10-174-582-478	Sequence 478, App

20	2249	70.1	800	US-10-174-588-478	Sequence 478, App
21	2249	70.1	800	US-10-175-739-478	Sequence 478, App
22	2249	70.1	800	US-10-175-740-478	Sequence 478, App
23	2249	70.1	800	US-10-175-743-478	Sequence 478, App
24	2249	70.1	800	US-10-176-488-478	Sequence 478, App
25	2249	70.1	800	US-10-176-492-478	Sequence 478, App
26	2249	70.1	800	US-10-176-747-478	Sequence 478, App
27	2249	70.1	800	US-10-176-750-478	Sequence 478, App
28	2249	70.1	800	US-10-176-985-478	Sequence 478, App
29	2249	70.1	800	US-10-176-985-478	Sequence 478, App
30	2249	70.1	800	US-10-176-991-478	Sequence 478, App
31	2249	70.1	800	US-10-176-992-478	Sequence 478, App
32	2249	70.1	800	US-10-176-993-478	Sequence 478, App
33	2249	70.1	800	US-10-184-658-478	Sequence 478, App
34	2249	70.1	800	US-10-036-214-52	Sequence 52, App1
35	2249	70.1	800	US-10-173-695-478	Sequence 478, App
36	2249	70.1	800	US-10-173-697-478	Sequence 478, App
37	2249	70.1	800	US-10-173-705-478	Sequence 478, App
38	2249	70.1	800	US-10-174-576-478	Sequence 478, App
39	2249	70.1	800	US-10-174-585-478	Sequence 478, App
40	2249	70.1	800	US-10-174-586-478	Sequence 478, App
41	2249	70.1	800	US-10-175-747-478	Sequence 478, App
42	2249	70.1	800	US-10-176-481-478	Sequence 478, App
43	2249	70.1	800	US-10-176-485-478	Sequence 478, App
44	2249	70.1	800	US-10-176-487-478	Sequence 478, App
45	2249	70.1	800	US-10-176-493-478	Sequence 478, App

#### ALIGNMENTS

RESULT 1  
US-10-050-704-125  
Sequence 125, Application US/10050704  
Publication No. US20030050442A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et al.  
TITLE OF INVENTION: 62 Human Secreted Proteins  
FILE REFERENCE: P2039P1  
CURRENT APPLICATION NUMBER: US/10/050,704  
CURRENT FILING DATE: 2002-01-18  
PRIOR APPLICATION NUMBER: 09/684,524  
PRIOR FILING DATE: 2000-10-10  
PRIOR APPLICATION NUMBER: PCT/US00/08979  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/128,693  
PRIOR FILING DATE: 1999-04-09  
PRIOR APPLICATION NUMBER: 60/130,991  
PRIOR FILING DATE: 1999-04-26  
NUMBER OF SEQ ID NOS: 344  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 125  
LENGTH: 627  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-050-704-125

Query Match 100.0%; Score 3208; DB 9; Length 627;  
Best Local Similarity 100.0%; Pred. No. 1.6e-228;  
Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEARVHALQKROVSLICVFLGVSMAEPLRYVAEETKGFPLAMIAIDLGIVEL 60  
DB 1 MEARVHALQKROVSLICVFLGVSMAEPLRYVAEETKGFPLAMIAIDLGIVEL 60  
QY 61 ARGRIVSDEFTIGLLNPLTGDLINLEKDKRELCEGTEPCVLPFOLLLEKPPQIRAE 120  
DB 61 ARGRIVSDEFTIGLLNPLTGDLINLEKDKRELCEGTEPCVLPFOLLLEKPPQIRAE 120  
QY 121 LMRVDINDSHSVFLDRETTINLESTPGATFLESADSDVGINNLRNTTISNVYFHI 180  
DB 121 LMRVDINDSHSVFLDRETTINLESTPGATFLESADSDVGINNLRNTTISNVYFHI 180



QY 181 NVHNDGEGNVSELYLVLDKVLDEEVEPELRLTLTGDDGSPRSGTTLRIILVLDINDVP 240  
DB 181 NVHNDGEGNVSELYLVLDKVLDEEVEPELRLTLTGDDGSPRSGTTLRIILVLDINDVP 240  
QY 241 EEVESLYKVQVPEVPSVSLVYVTSARDLDTGSGNEIYFAFEYATEERTLKTFRINSTGN 300  
DB 241 EEVESLYKVQVPEVPSVSLVYVTSARDLDTGSGNEIYFAFEYATEERTLKTFRINSTGN 300  
QY 301 LHLKELVYEAIOYITLTITQAKDGGSLGKCTVVVHVTDINDNPELIMSSLTSPIDENS 360  
DB 301 LHLKELVYEAIOYITLTITQAKDGGSLGKCTVVVHVTDINDNPELIMSSLTSPIDENS 360  
QY 361 PETVAVERIRDRDGGNNAKWCSTIOHLPVLKPSVNEFTVLTERRADDEEETENIT 420  
DB 361 PETVAVERIRDRDGGNNAKWCSTIOHLPVLKPSVNEFTVLTERRADDEEETENIT 420  
QY 421 ITVTDLGTPRLKTOHNLVTVSDVNDNAPEFSQTTTLRVRENNSPALHIGSVATDRDS 480  
DB 421 ITVTDLGTPRLKTOHNLVTVSDVNDNAPEFSQTTTLRVRENNSPALHIGSVATDRDS 480  
QY 481 GANAQVTSLLPPHDPQLPLGSLVSINADNGOLFALNSLDFEALQAEFRVGAADRSPA 540  
DB 481 GANAQVTSLLPPHDPQLPLGSLVSINADNGOLFALNSLDFEALQAEFRVGAADRSPA 540  
QY 541 LSSQALVYLVADANDNAPEVLYPLONGSARCTELVPRAAGYLVAKVAVDGDSCQA 600  
DB 541 LSSQALVYLVADANDNAPEVLYPLONGSARCTELVPRAAGYLVAKVAVDGDSCQA 600  
QY 601 WLSTYOLKATEPGLGFWAHNGEVRTA 627  
DB 601 WLSTYOLKATEPGLGFWAHNGEVRTA 627

RESULT 2  
US-10-036-041-52  
Sequence 52, Application US/10036041  
Publication No. US20020192751A1  
GENERAL INFORMATION:  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3030R1C8  
CURRENT APPLICATION NUMBER: US/10/036,041  
PRIOR FILING DATE: 2001-12-26  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/112514  
PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: 60/113300  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/113430  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113605  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113621  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/114140  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/115552  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/116843  
PRIOR FILING DATE: 1999-01-22  
PRIOR APPLICATION NUMBER: 60/125774  
PRIOR FILING DATE: 1999-03-23

PRIOR APPLICATION NUMBER: 60/125778  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: 60/125826  
PRIOR FILING DATE: 1999-03-24  
PRIOR APPLICATION NUMBER: 60/127035  
PRIOR FILING DATE: 1999-03-31  
PRIOR APPLICATION NUMBER: 60/127706  
PRIOR FILING DATE: 1999-04-05  
PRIOR APPLICATION NUMBER: 60/129122  
PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: 60/130359  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: 60/131270  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131272  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131291  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/132371  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/132379  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/132383  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/135750  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: 60/138166  
PRIOR FILING DATE: 1999-06-08  
PRIOR APPLICATION NUMBER: 60/144791  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/146970  
PRIOR FILING DATE: 1999-08-03  
PRIOR APPLICATION NUMBER: 60/162506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: 09/311832  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 09/380142  
PRIOR FILING DATE: 1999-08-25  
PRIOR APPLICATION NUMBER: 09/644848  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 09/747259  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 09/816744  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 09/854208  
PRIOR FILING DATE: 2001-05-10  
PRIOR APPLICATION NUMBER: 09/854280  
PRIOR FILING DATE: 2001-05-10  
PRIOR APPLICATION NUMBER: 09/874503  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: 09/869599  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: 09/908,827  
PRIOR FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: PCT/US99/10733  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: PCT/US99/28551  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30720  
PRIOR FILING DATE: 1999-12-22  
PRIOR APPLICATION NUMBER: PCT/US00/05601  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: PCT/US00/05841  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: PCT/US00/14042  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: PCT/US00/15264  
PRIOR FILING DATE: 2000-06-02  
PRIOR APPLICATION NUMBER: PCT/US00/23522  
PRIOR FILING DATE: 2000-08-23  
PRIOR APPLICATION NUMBER: PCT/US00/23328  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: PCT/US00/32678

```

PRT      PRIOR FILING DATE: 2000-12-01
PRT      PRIOR APPLICATION NUMBER: PCT/US00/034956
PRT      PRIOR FILING DATE: 2000-12-20
PRT      PRIOR APPLICATION NUMBER: PCT/US01/06520
PRT      PRIOR FILING DATE: 2001-02-28
PRT      PRIOR APPLICATION NUMBER: PCT/US01/17800
PRT      PRIOR FILING DATE: 2001-06-01
PRT      PRIOR APPLICATION NUMBER: PCT/US01/19692
PRT      PRIOR FILING DATE: 2001-06-20
PRT      PRIOR APPLICATION NUMBER: PCT/US01/21066
PRT      PRIOR FILING DATE: 2001-06-29
PRT      PRIOR APPLICATION NUMBER: PCT/US01/21735
PRT      PRIOR FILING DATE: 2001-07-09
PRT      NUMBER OF SEQ ID NOS: 80
PRT      SEQ ID NO 52
PRT      LENGTH: 800
PRT      TYPE: PRT
PRT      ORGANISM: Homo Sapien
PRT      US-10-036-041-52

```

```

? PRIOR FILING DATE: 1999-08-25
? PRIOR APPLICATION NUMBER: 09/644848
? PRIOR FILING DATE: 2000-08-22
? PRIOR APPLICATION NUMBER: 09/747259
? PRIOR FILING DATE: 2000-12-20
? PRIOR APPLICATION NUMBER: 09/816744
? PRIOR FILING DATE: 2001-03-22
? PRIOR APPLICATION NUMBER: 09/854208
? PRIOR FILING DATE: 2001-05-10
? PRIOR APPLICATION NUMBER: 09/854280
? PRIOR FILING DATE: 2001-05-10
? PRIOR APPLICATION NUMBER: 09/874503
? PRIOR FILING DATE: 2001-06-05
? PRIOR APPLICATION NUMBER: 09/869599
? PRIOR FILING DATE: 2001-06-29
? PRIOR APPLICATION NUMBER: 09/908, 827
? PRIOR FILING DATE: 2001-07-18
? PRIOR APPLICATION NUMBER: PCT/US99/10733
? PRIOR FILING DATE: 1999-05-14
? PRIOR APPLICATION NUMBER: PCT/US99/28551
? PRIOR FILING DATE: 1999-12-02
? PRIOR APPLICATION NUMBER: PCT/US99/30720
? PRIOR FILING DATE: 1999-12-22
? PRIOR APPLICATION NUMBER: PCT/US00/05601
? PRIOR FILING DATE: 2000-03-01
? PRIOR APPLICATION NUMBER: PCT/US00/05841
? PRIOR FILING DATE: 2000-03-02
? PRIOR APPLICATION NUMBER: PCT/US00/14042
? PRIOR FILING DATE: 2000-05-22
? PRIOR APPLICATION NUMBER: PCT/US00/15264
? PRIOR FILING DATE: 2000-06-02
? PRIOR APPLICATION NUMBER: PCT/US00/23522
? PRIOR FILING DATE: 2000-08-23
? PRIOR APPLICATION NUMBER: PCT/US00/23328
? PRIOR FILING DATE: 2000-08-24
? PRIOR APPLICATION NUMBER: PCT/US00/32678
? PRIOR FILING DATE: 2000-12-01
? PRIOR APPLICATION NUMBER: PCT/US00/34956
? PRIOR FILING DATE: 2000-12-20
? PRIOR APPLICATION NUMBER: PCT/US01/06520
? PRIOR FILING DATE: 2001-02-28
? PRIOR APPLICATION NUMBER: PCT/US01/17800
? PRIOR FILING DATE: 2001-06-01
? PRIOR APPLICATION NUMBER: PCT/US01/19692
? PRIOR FILING DATE: 2001-06-20
? PRIOR APPLICATION NUMBER: PCT/US01/21066
? PRIOR FILING DATE: 2001-06-29
? PRIOR APPLICATION NUMBER: PCT/US01/21735
? PRIOR FILING DATE: 2001-07-09
? NUMBER OF SEQ ID NOS: 80
? SEQ ID NO 52
? LENGTH: 800
? TYPE: PRT
? ORGANISM: Homo Sapien
US-10-035-855-52

```

```

Query Match 70.1%; Score 2249; DB 9; Length 800;
Best Local Similarity 70.4%; Pred. No. 1.2e-157;
Matches 435; Conservative 70; Mismatches 113; Indels 0; Gaps 0;

```

```

QY 10 OKROYSLICVFLGYSMAGAEPLRYFAEETRGFTFLANLADLGLVEELISARGCRIVSD 69
DB 10 ROROVLEFLFWMGVSILAGSGFGRYSVTEETKESFVNLAKDLGLAEGELANGSTRVSD 69
QY 70 ETIFLLINPLTGDLLNEKLDREELCGPTEPCVLPOLLLEKFPQIFRAELNVRDINDH 129
DB 70 DNKOYLLDLSHTGNLTNEKLDREELCGPTEPCVLPOLLLEKFPQIFRAELNVRDINDH 129
QY 130 SPVFLDEITLNTLESTTPGATEFLLESABSDVGINNLNRYTSSNVYFHHVNDGEGN 189
DB 130 APVPODEYVLTAKISENTAECTAFRLERADPDGGLNGCIOWYTTSPNSFHHINISGDEGM 189
QY 190 VYSELVLDKVLDEEYVELRLTLTGIDGSPPSGFTLLIRIILVDINDNVPFVESLYKY 249

```

```

DB 190 IYPELVLDKALDREOQELSLTILALDGGSPSSGTSVRIIVLDVNDNAPOQAOLYER 249
QY 250 QVNPSPVGSIVYVYVSARDIDTGSNGRIYAFYATFRTKTRINSTSGNLHKLKELNY 309
DB 250 QAPENSPITGLYKVMWEDVDGVAEVSYSFFDASINRTTQIINFSGEFLRELDLY 309
QY 310 EAIOTYTLTIOAKDGGLSGKCTVWVHDINDNPELMSLTSPIPENSPETVAVR 369
DB 310 ELVYSKINIOAMDGGLSARCRIVLEVDTNPNPELLVSSISNVANSPEPTLAVRK 369
QY 370 IRRDSSNNAKMCISIDDLPELVKPSVENFTLVTERALDREREYNTITVYDLGRP 429
DB 370 INRDSSENGKWCYIOENLPLFKPSVENFTLITEGALDREREYNTITVYDLGRP 429
QY 430 RLKQHLNLTYSVDVNDNAFTFSOTYTLTREVNSPALHIGSVSTDDSDGNAQYTS 489
DB 430 RLKTEHNTYLVSDVNDNAFTFSOTYTLTREVNSPALHIGSVSTDDSDGNAQYTS 489
QY 490 LPPHDPQLPLGSLVSLNADNGQLFALRLDFAALQAFERVGADRGSPALSGQALVRY 549
DB 490 LPPHDPQLPLGSLVSLNADNGQLFALRLDFAALQAFERVGADRGSPALSGQALVRY 549
QY 550 LVADANDNAFVLYXPLONGSAPCTELVPPRAEAGLYVAVVAVDGDSCGNMLSTQLKA 609
DB 550 LVADANDNSFVLYPLONGSAPCTELVPPRAEAGLYVAVVAVDGDSCGNMLSTQLKA 609
QY 610 TERGLFGVMAHNGEVRTA 627
DB 610 TERGLFGVMAHNGEVRTA 627

```

```

RESULT 4
US-10-174-590-478
? Sequence 478, Application US/10174590
? Publication No. US20030008352A1
? GENERAL INFORMATION:
? APPLICANT: Baker, Kevin P.
? APPLICANT: Chen, Jian
? APPLICANT: Desnoyers, Luc
? APPLICANT: Goddard, Audrey
? APPLICANT: Godowski, Paul J.
? APPLICANT: Gurney, Austin L.
? APPLICANT: Pan, James
? APPLICANT: Smith, Victoria
? APPLICANT: Watanabe, Colin K.
? APPLICANT: Wood, William I.
? APPLICANT: Zhang, Zemin
? TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
? FILE REFERENCE: P3430R1C42
? CURRENT APPLICATION NUMBER: US/10/174, 590
? PRIOR FILING DATE: 2002-06-18
? PRIOR APPLICATION removed - See file Wrapper or Palm
? NUMBER OF SEQ ID NOS: 612
? SEQ ID NO 478
? LENGTH: 800
? TYPE: PRT
? ORGANISM: Homo Sapien
US-10-174-590-478

```

```

Query Match 70.1%; Score 2249; DB 9; Length 800;
Best Local Similarity 70.4%; Pred. No. 1.2e-157;
Matches 435; Conservative 70; Mismatches 113; Indels 0; Gaps 0;

```

```

QY 10 OKROYSLICVFLGYSMAGAEPLRYFAEETRGFTFLANLADLGLVEELISARGCRIVSD 69
DB 10 ROROVLEFLFWMGVSILAGSGFGRYSVTEETKESFVNLAKDLGLAEGELANGSTRVSD 69
QY 70 ETIFLLINPLTGDLLNEKLDREELCGPTEPCVLPOLLLEKFPQIFRAELNVRDINDH 129
DB 70 DNKOYLLDLSHTGNLTNEKLDREELCGPTEPCVLPOLLLEKFPQIFRAELNVRDINDH 129

```

	Query Match	70.1%	Score	2249:	DB	9:	Length	800:
	Best Local Similarity	70.4%:	Pred.	No.	1,2e-157:			
	Matches	435:	Conservative	70:	Mismatches	113:	Indels	0:
	Gaps							0:
QY	10	OKROYSLICVFLGVSAGACPELRIFYAEETENGTFPLANAIADLGLGVFELSARCRIVSD	69					
		:::::						
DB	10	RROVAFLFLPMGVSLAGSGFGRIYSTETETEGSFVNVIANKIDJLAEGGLAARGTIVSD	69					
		:::						

Query Match	70.1%;	Score 2249;	DB 9;	Length 800;
Best Local Similarity	70.4%;	Pred. No. 1.2e-157;		

[illegible][illegible]

REFUGANT: MOO, William T.  
; APPLICANT: Zhang, Zemin

APPLICANT: Gurney, Austin L.

```

; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C7D
; CURRENT APPLICATION NUMBER: US/10/176,482
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 478
; LENGTH: 800
; TYPE: PRP
; ORGANISM: Homo Sapien
US-10-176-482-478

```

```

Query Match 70.1%; Score 2249; DB 9; Length 800;
Best Local Similarity 70.4%; Pred. NO. 1.2e-157;
Matches 435; Conservative 70; Mismatches 113; Indels 0; Gaps 0;

```

```

QY 10 OKROYSLICVPLGVSMAGAEPLRYFAETERTGTFLANLADIGLVEELSGRCRIYSD 69
Db 10 RORQVLFLEFMGYSLAGSGFRYSVTEETKSGFVNKLADGLAGELAAKGRIVSD 69
QY 70 ETIGFLNPLTGDLNLEKLDREELCGPTPECVLPQILLKPKQIFRAELWVDINDH 129
Db 70 DNKOYLLDLSHTGNLITNEKLDREELCGPKPECMLYFOIIMDDPQIYRAELRVARDINDH 129
QY 130 SPVELDREITINLESTPGATFLLSAGSDVGINNLNRTTSSNVEFHINVDNCEGN 189
Db 130 APVQDKETVIAKISENTAEGTAFLERAODPDGGLNGIQNTTSPNSFHHINISGDEGM 189
QY 190 VYSELVLDKVDLREVEBELRLTLTGDLGSPRSSTLIRILYLDINDNPEFEVESLYKV 249
Db 190 IYPELVLDKALDREOGEELSLTLTGDLGSPRSSTVRLVLDVNDNAPOFAOLYET 249
QY 250 QVPENSPVGLVYVSARDLDGTSNGEIVYAFYATERTKTRINSTSGNLHKAELNY 309
Db 250 QAPENSPIGFLIVKVAEDVDGVAEVSYSFEDASENIRTTFOINPFSGEIFLRELDY 309
QY 310 EAIOTYTLTIQAKDGGGLSGKCTVYVHTDINDNPELLMSLSPIPEPSPETVAVFR 369
Db 310 ELVNSYKINTIQAMGGGLSARCRVLYEVLDTNDNPELLVSSFSNSVAENSPETPLAVFR 369
QY 370 IRODPSGNNAKWCISIQDHLPEVLKPSVENFTLYTERALDREERTYNTITVTDLGP 429
Db 370 INDRDSENGKMKVCYIOENLPFLKPSVENFTLITGALDREIRAEVNTITVTDLGP 429
QY 430 RLKTOHNLTVYSDVNDNAPEFTSQTTTLRVRENNSPALHIGVSATDRDSCANAQVYS 489
Db 430 RLKTEHNTVYSDVNDNAPEFTQSTTLRVRENNSPALHIGVSATDRDSCGNAQVYS 489
QY 490 LLPHDPOLPLGSLVSIINADNGQLFALRSIDPEALQAEFRVGAAGSPALSSQALVRY 549
Db 490 LLPPQDPLPLASLVSINADNGHLFALRSIDYELALQAEFRVGAATRGSPALSSREALVRY 549
QY 550 LVADANDNAPFVLYPLONGSAPCTELVPRAAEAGYLVAKVVAVDGSGONAMLSYOLKA 609
Db 550 LVADANDNSPFVLYPLONGSAPCTELVPRAAEAGYLVAKVVAVDGSGONAMLSYOLKA 609
QY 610 TERGLFGVMAHNGEVRTA 627
Db 610 TERGLFGVMAHNGEVRTA 627

```

```

RESULT 11
US-10-176-757-478
; Sequence 478, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

```

```

; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C86
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 478
; LENGTH: 800
; TYPE: PRP
; ORGANISM: Homo Sapien
US-10-176-757-478

```

```

Query Match 70.1%; Score 2249; DB 9; Length 800;
Best Local Similarity 70.4%; Pred. NO. 1.2e-157;
Matches 435; Conservative 70; Mismatches 113; Indels 0; Gaps 0;

```

```

QY 10 OKROYSLICVPLGVSMAGAEPLRYFAETERTGTFLANLADIGLVEELSGRCRIYSD 69
Db 10 RORQVLFLEFMGYSLAGSGFRYSVTEETKSGFVNKLADGLAGELAAKGRIVSD 69
QY 70 ETIGFLNPLTGDLNLEKLDREELCGPTPECVLPQILLKPKQIFRAELWVDINDH 129
Db 70 DNKOYLLDLSHTGNLITNEKLDREELCGPKPECMLYFOIIMDDPQIYRAELRVARDINDH 129
QY 130 SPVELDREITINLESTPGATFLLSAGSDVGINNLNRTTSSNVEFHINVDNCEGN 189
Db 130 APVQDKETVIAKISENTAEGTAFLERAODPDGGLNGIQNTTSPNSFHHINISGDEGM 189
QY 190 VYSELVLDKVDLREVEBELRLTLTGDLGSPRSSTLIRILYLDINDNPEFEVESLYKV 249
Db 190 IYPELVLDKALDREOGEELSLTLTGDLGSPRSSTVRLVLDVNDNAPOFAOLYET 249
QY 250 QVPENSPVGLVYVSARDLDGTSNGEIVYAFYATERTKTRINSTSGNLHKAELNY 309
Db 250 QAPENSPIGFLIVKVAEDVDGVAEVSYSFEDASENIRTTFOINPFSGEIFLRELDY 309
QY 310 EAIOTYTLTIQAKDGGGLSGKCTVYVHTDINDNPELLMSLSPIPEPSPETVAVFR 369
Db 310 ELVNSYKINTIQAMGGGLSARCRVLYEVLDTNDNPELLVSSFSNSVAENSPETPLAVFR 369
QY 370 IRODPSGNNAKWCISIQDHLPEVLKPSVENFTLYTERALDREERTYNTITVTDLGP 429
Db 370 INDRDSENGKMKVCYIOENLPFLKPSVENFTLITGALDREIRAEVNTITVTDLGP 429
QY 430 RLKTOHNLTVYSDVNDNAPEFTSQTTTLRVRENNSPALHIGVSATDRDSCANAQVYS 489
Db 430 RLKTEHNTVYSDVNDNAPEFTQSTTLRVRENNSPALHIGVSATDRDSCGNAQVYS 489
QY 490 LLPHDPOLPLGSLVSIINADNGQLFALRSIDPEALQAEFRVGAAGSPALSSQALVRY 549
Db 490 LLPPQDPLPLASLVSINADNGHLFALRSIDYELALQAEFRVGAATRGSPALSSREALVRY 549
QY 550 LVADANDNAPFVLYPLONGSAPCTELVPRAAEAGYLVAKVVAVDGSGONAMLSYOLKA 609
Db 550 LVADANDNSPFVLYPLONGSAPCTELVPRAAEAGYLVAKVVAVDGSGONAMLSYOLKA 609
QY 610 TERGLFGVMAHNGEVRTA 627
Db 610 TERGLFGVMAHNGEVRTA 627

```

RESULT 12





QY 550 LVADANDNAPFVLYPLONGSAPCTELVPRAAEGYLVAKVAVDGDSCGNAMLSYOLLLKA 609  
DB 550 LVADANDNAPFVLYPLONGSAPCTELVPRAAEGYLVAKVAVDGDSCGNAMLSYOLLLKA 609  
QY 610 TERPGLFGVWAHNGEVRTA 627  
DB 610 TERPGLFGVWAHNGEVRTA 627

## RESULT 14

US-10-180-557-478  
Sequence 478, Application US/10180557  
Publication No. US20030022301A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C147  
CURRENT APPLICATION NUMBER: US/10/180,557  
PRIORITY FILING DATE: 2002-06-25  
Prior Application removed - See file Wrapper or Palm  
SEQU ID NO 478  
LENGTH: 800  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-180-557-478

## Query Match

Best Local Similarity 70.1%; Score 2249; DB 9; Length 800;  
Matches 435; Conservative 70; Mismatches 113; Indels 0; Gaps 0;

QY 10 QKROVSLCYFLGVSWAGAPLRFVAETERGFTLANLIDGLGVEELASAGCTIVSD 69  
DB 10 RQROVLFLEFWGSLAGSGFRKSYTEEEKSFVNLAKDKGLGELAAAGTAVSD 69  
QY 70 ETIGELLNPLDGLLNEKLDREELCGPTEPCVLPQDLLEKPFQIFRAELWVRDINDH 129  
DB 70 DNKQYLLDSDHTGNLFTNEKLDREELCGPTEPCVLPQDLLEKPFQIFRAELWVRDINDH 129  
QY 130 SPVFLDEITLNLIESTTPGATFLLESADSDVGINNENYTTSSNVYPHINVDNGEEN 189  
DB 130 APVFODEYVLKISENTEAETARLEAODPDGLNGIQNYTISPNSFPHINISGDEGN 189  
QY 190 VSELVADKVLDEEVEVELRLTLGLDGGSPRSGTTLNLIYLDNDNPEVESESLXV 249  
DB 190 IYELVADKVLDEEVEVELRLTLGLDGGSPRSGTTLNLIYLDNDNPEVESESLXV 249  
QY 250 QVBNSEVSVLVTVSARDIDTGSNEIYVAFYATERTLKTFRINSGNLTKAELNY 309  
DB 250 QVBNSEVSVLVTVSARDIDTGSNEIYVAFYATERTLKTFRINSGNLTKAELNY 309  
QY 310 ELAIOITLTIOAKDGLSGKCTVYVHYVDINDNPELMSLSLTPENSPETVAVVR 369  
DB 310 ELVNSYKINIQAMDGGLSARCVLEVDINDNPELMSLSLTPENSPETVAVVR 369  
QY 370 IRDRSGNNAKVCSIODHLPFLKPSVENYTLVTERALDREPEREYNTITVVDLGR 429  
DB 370 INDRSGNNAKVCSIODHLPFLKPSVENYTLVTERALDREPEREYNTITVVDLGR 429  
QY 430 RLKTOHNLVTVSDVNDNAPFESQTYTLVRENNSPALHISVSATDSDGANAQVYS 489  
DB 430 RLKTEHNTVTVSDVNDNAPFESQTYTLVRENNSPALHISVSATDSDGANAQVYS 489

QY 490 LPPHDPPLIGSLVAINDNQGLFALRSIDFEALQAEFRYRGADRGSALSSQALVRY 549  
DB 490 LPPHDPPLIGSLVAINDNQGLFALRSIDFEALQAEFRYRGADRGSALSSQALVRY 549  
QY 550 LVADANDNAPFVLYPLONGSAPCTELVPRAAEGYLVAKVAVDGDSCGNAMLSYOLLLKA 609  
DB 550 LVADANDNAPFVLYPLONGSAPCTELVPRAAEGYLVAKVAVDGDSCGNAMLSYOLLLKA 609  
QY 610 TERPGLFGVWAHNGEVRTA 627  
DB 610 TERPGLFGVWAHNGEVRTA 627

## RESULT 15

US-09-931-836-52  
Sequence 52, Application US/09931836  
Publication No. US2003002749A1  
GENERAL INFORMATION:  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3030R1C1  
CURRENT APPLICATION NUMBER: US/09/931,836  
PRIORITY FILING DATE: 2001-08-16  
Prior Application removed - See file Wrapper or Palm  
SEQU ID NO 478  
LENGTH: 800  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-09-931-836-52

PRIOR FILING DATE	1999-05-04
PRIOR APPLICATION NUMBER:	60/1132379
PRIOR FILING DATE:	1999-05-04
PRIOR APPLICATION NUMBER:	60/113283
PRIOR FILING DATE:	1999-05-04
PRIOR APPLICATION NUMBER:	60/135750
PRIOR FILING DATE:	1999-05-25
PRIOR APPLICATION NUMBER:	60/138166
PRIOR FILING DATE:	1999-06-08
PRIOR APPLICATION NUMBER:	60/144791
PRIOR FILING DATE:	1999-07-20
PRIOR APPLICATION NUMBER:	60/146970
PRIOR FILING DATE:	1999-08-03
PRIOR APPLICATION NUMBER:	60/162506
PRIOR FILING DATE:	1999-10-29
PRIOR APPLICATION NUMBER:	09/311832
PRIOR FILING DATE:	1999-05-14
PRIOR APPLICATION NUMBER:	09/380142
PRIOR FILING DATE:	1999-08-25
PRIOR APPLICATION NUMBER:	09/644848
PRIOR FILING DATE:	2000-08-82
PRIOR APPLICATION NUMBER:	09/747259
PRIOR FILING DATE:	2000-12-20
PRIOR APPLICATION NUMBER:	09/816744
PRIOR FILING DATE:	2001-03-82
PRIOR APPLICATION NUMBER:	09/854208
PRIOR FILING DATE:	2001-05-10
PRIOR APPLICATION NUMBER:	09/854280
PRIOR FILING DATE:	2001-05-10
PRIOR APPLICATION NUMBER:	09/874503
PRIOR FILING DATE:	2001-06-05
PRIOR APPLICATION NUMBER:	09/869599
PRIOR FILING DATE:	2001-06-29
PRIOR APPLICATION NUMBER:	09/908_827
PRIOR FILING DATE:	2001-07-18
PRIOR APPLICATION NUMBER:	PCT/US99/107333
PRIOR FILING DATE:	1999-05-14
PRIOR APPLICATION NUMBER:	PCT/US99/285513
PRIOR FILING DATE:	1999-12-02
PRIOR APPLICATION NUMBER:	PCT/US99/307200
PRIOR FILING DATE:	1999-12-22
PRIOR APPLICATION NUMBER:	PCT/US00/056011
PRIOR FILING DATE:	2000-03-01
PRIOR APPLICATION NUMBER:	PCT/US00/058411
PRIOR FILING DATE:	2000-03-02
PRIOR APPLICATION NUMBER:	PCT/US00/140422
PRIOR FILING DATE:	2000-05-22
PRIOR APPLICATION NUMBER:	PCT/US00/152664
PRIOR FILING DATE:	2000-06-02
PRIOR APPLICATION NUMBER:	PCT/US00/235222
PRIOR FILING DATE:	2000-08-23
PRIOR APPLICATION NUMBER:	PCT/US00/233288
PRIOR FILING DATE:	2000-08-24
PRIOR APPLICATION NUMBER:	PCT/US00/326787
PRIOR FILING DATE:	2000-12-01
PRIOR APPLICATION NUMBER:	PCT/US00/349556
PRIOR FILING DATE:	2000-12-20
PRIOR APPLICATION NUMBER:	PCT/US01/065200
PRIOR FILING DATE:	2001-08-28
PRIOR APPLICATION NUMBER:	PCT/US01/178000
PRIOR FILING DATE:	2001-06-01
PRIOR APPLICATION NUMBER:	PCT/US01/196922
PRIOR FILING DATE:	2001-06-20
PRIOR APPLICATION NUMBER:	PCT/US01/210666
PRIOR FILING DATE:	2001-06-29
PRIOR APPLICATION NUMBER:	PCT/US01/217355
PRIOR FILING DATE:	2001-07-09
NUMBER OF SEQ ID NOS:	80
SEQ ID NO	52
LENGTH:	800
TYPE:	PRT
ORGANISM:	Homo Sapien
OS:	09-931-836-52

Query Match	70.1%;	Score 2249;	DB 9;	Length 800;
Best Local Similarity	70.4%;	Pred. NO. 1.2e-157;		
Matches 435;	Conservative 70;	Mismatches 113;	Indels 0;	Gaps 0;

QY	10	OKROVSLICVYASMAKGADELFVFAKEERGFPLANLADLGVLGEVLSAGCRIVSD	69
Db	10	KOROVLEFLFELWGVSLAGSGEGKRVSYEEFEKGSFVYNLAKDGLAGELAAAGTRIVSD	69
QY	70	ETIGFLLNPLTGDLLNEKIDREELCGPTPECVLPQELLEKRFQIFRELWVRDINH	129
Db	70	DNKQYLLDSHTGMLTNEKIDREKLCGPKPECKMLYQIIMDDPFOYRAELRVDRINDH	129
QY	130	SPVFLDRITLITLIESTTPATFLLESAHSDVGINLBRXTITSSNVPFIINHNDGECN	189
Db	130	APVEQDKETVLKISEENRAGFAERLEAOPDDGLNIOYTTISPNSFPHINISGDEGN	189
QY	190	VYSELVLDKVLDRREVEBELRTLTGLDGGSPRSCTLLIRILVLDINDNVPFVESILYKV	249
Db	190	IYPELVLDKALDRREGSLSTLTALDGGSPRSRGCTVRIVLDVNDNAPFOALYER	249
QY	250	QVPENSPVGSLVTVSAROLDTGSNGELIYVAFYATERTLKTIRINSTGNLHKLELNY	305
Db	250	QAPENSPIGFLIVYMAEDVDGSNAEVSYSFEDPASNIRITQIINFSEIIFRELXY	309
QY	310	BAIORYTTLIOAKOGGGLSGCTVYVYVHTDINDNPELLMSSLTSPJEPKSPETVAVER	369
Db	310	EIVNYSKATINIOAMOGGILSACRCVLVEYLDINDNPPELIYSSGSNSNAEKSPEPLAVKR	369
QY	370	IRDRDSGNNAKWCYSIODHLPFLVKPSVENEYTLVTERALDREREYNTITVTDLGNP	429
Db	370	INDRDSGENKMWCIYOENLPFLKPSVENEYTLITGALDRITRAEYNTITVTDLGNP	429
QY	430	RLKTOHNTLVYSDVNDNAPFQSTQTYTLRRENNSPALHIGVSADRDROGNAOVTYS	489
Db	430	RLKTEHNTLVYSDVNDNAPFQSTQTYTLRRENNSPALHIGVSADRDROGNAOVTYS	489
QY	490	LLPFPDPLPLGSLVSIINADNGOLFALRSLDLFEALQAFEEFVCGAADRGSPALSSQALVRV	549
Db	490	LLPDPDPLPLASLVSINADNGHFLALRSLDYELALQAFEEFVCGATIDGSPALSSREALVRV	549
QY	550	LVADANDNAPVLVYPLONGSAPCTELVPRAAEPCYLVTVKVAVDGDSGONAMLSYOLLA	609
Db	550	LVLDANDNSPEVLVPLONGSAPCTELVPRAAEPCYLVTVKVAVDGDSGONAMLSYOLLA	609
QY	610	TEPGEGFVYVWANGVEYRTA	627
Db	610	TEPGLEFVYVWANGVEYRTA	627

Search completed: May 16, 2003, 10:48:51  
Job time : 27 secs



GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 10:45:57 ; Search time 14 Seconds  
(Without alignments)  
1857.547 million cell updates/sec

Title: US-10-050-704-125

Perfect score: 3208

Sequence: 1 MEARVHALQKQVSLICVF.....KATPGLFVYVHNGEVRTA 627

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2752	85.8	793	1	CDB7_HUMAN
2	2323	72.4	798	1	CDB2_HUMAN
3	2306	71.9	798	1	CDBE_HUMAN
4	2287	71.3	787	1	CDBF_HUMAN
5	2262	70.5	776	1	CDBG_HUMAN
6	2250	70.1	776	1	CDBH_HUMAN
7	2249	70.1	797	1	CDBA_HUMAN
8	2247	70.0	797	1	CDB9_HUMAN
9	2205	68.7	795	1	CDB4_HUMAN
10	2181	68.0	795	1	CDBC_HUMAN
11	2179.5	67.9	795	1	CDB5_HUMAN
12	2174.5	67.8	794	1	CDB6_HUMAN
13	2171	67.7	797	1	CDB8_HUMAN
14	2156.5	67.2	801	1	CDBD_HUMAN
15	2142	66.8	798	1	CDBH_HUMAN
16	2009.5	62.6	797	1	CDB3_HUMAN
17	1749.5	59.4	818	1	CDB1_HUMAN
18	1264.5	39.4	950	1	CDA3_HUMAN
19	1239.5	38.6	936	1	CDA5_HUMAN
20	1234.5	38.5	1007	1	CHC2_HUMAN
21	1230.5	38.4	950	1	CDA2_HUMAN
22	1227	38.2	948	1	CDA1_HUMAN
23	1222	38.1	950	1	CDA1_HUMAN
24	1203	37.5	937	1	CDA7_HUMAN
25	1200.5	37.4	947	1	CDA4_HUMAN
26	1200	37.4	950	1	CDA6_HUMAN
27	1194	37.2	950	1	CDA6_HUMAN
28	1192.5	37.2	949	1	CDA6_HUMAN
29	1188	37.0	963	1	CHC1_HUMAN
30	1179	36.8	941	1	CDA1_HUMAN
31	1175	36.6	950	1	CDA9_HUMAN
32	1157	36.1	948	1	CDA4_HUMAN
33	900.5	28.1	1180	1	PC12_MOUSE

34	880.5	27.4	1184	1	PC12_HUMAN	09npg4 homo sapien
35	807	25.2	1069	1	PCD7_HUMAN	060245 homo sapien
36	703	21.9	3298	1	PC16_HUMAN	096190 homo sapien
37	603	18.8	3313	1	CLR3_RAT	088278 rattus norv
38	599	18.7	3301	1	CLR3_MOUSE	091410 mus muscu
39	598.5	18.7	3312	1	CLR3_HUMAN	09nyg7 homo sapien
40	597	18.6	5147	1	FAT_DROME	P33450 drosophila
41	595.5	18.6	2923	1	CLR2_HUMAN	09hcu4 homo sapien
42	590	18.4	3014	1	CLR2_HUMAN	09nyg6 homo sapien
43	589	18.4	3579	1	STAN_DROME	09v5n8 drosophila
44	585	18.2	3354	1	CADN_HUMAN	09a251 homo sapien
45	584	18.2	3034	1	CLR1_MOUSE	035161 mus muscu

## ALIGNMENTS

RESULT 1  
CDB7\_HUMAN STANDARD; PRT; 793 AA.  
ID CDB7\_HUMAN  
AC Q9Y5E2;  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Protocadherin beta 7 precursor (PCDH-beta7).  
GN PCDH7.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=99308636; PubMed=10380929;  
RA Wu Q., Mandelis T.;  
RT "A striking organization of a large family of human neural cadherin-  
RL like cell adhesion genes.";  
RT Cell 97:779-790(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21223055; PubMed=11322959;  
RA Vanhaest K., Koels P., Vanden Eynde E., van Roy F.;  
RT "The human and murine protocadherin-beta one-exon gene families show  
RT high evolutionary conservation, despite the difference in gene  
RT number.";  
RL FEBS Lett. 495:120-125(2001).  
CC -!- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
CC -!- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
EMBL: AF152500; AAD43761.1; -;  
EMBL: AF217750; AAK51617.1; -;  
HSSP: P15116; INCU.  
DR GeneW; HGNC:8692; PCDH7.  
DR MIM; 606333; -;  
DR InterPro; IPR002126; Cadherin.  
DR Pfam; PF00205; CADHERIN.  
DR PRINTS; PRO0205; CADHERIN.  
DR SMART; SM00112; CA; 5.  
DR PROSITE; PS00332; CADHERIN\_1; 5.  
DR PROSITE; PS00268; CADHERIN\_2; 5.  
KW Calcium-binding; Cell adhesion; Glycoprotein; Signal; Repeat;  
KW Transmembrane; Multigene family.  
FT SIGNAL 1 26 POTENTIAL.  
FT CHAIN 27 793 PROTOCADHERIN BETA 7.

FT DOMAIN 27 688 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 689 709 POTENTIAL.  
 FT DOMAIN 710 793 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 35 133 CADHERIN 1.  
 FT DOMAIN 138 242 CADHERIN 2.  
 FT DOMAIN 247 347 CADHERIN 3.  
 FT DOMAIN 352 451 CADHERIN 4.  
 FT DOMAIN 456 561 CADHERIN 5.  
 FT DOMAIN 568 671 CADHERIN 6.  
 FT CARBOHYD 169 199 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 418 436 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 436 456 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 567 597 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 793 AA: 86706 MW: 85677098653C755 CRC64:

Query Match 85.8%; Score 2752; DB 1; Length 793;  
 Best Local Similarity 85.3%; Pred. No. 3.9e-16;  
 Matches 535; Conservative 34; Mismatches 58; Indels 0; Gaps 0;

QY 1 MEARVVALOKROVSLICVFLGVSWAGAPLRYFAEETGRTFLANLADLGVEELS 60  
 DB 1 MEARVVALOKROVSLICVFLGVSWAGAPLRYFAEETGRTFLANLADLGVEELS 60  
 QY 61 ARGRIVSDETIGFLLNPLDGLLNKDLREELCGTEPCVLPFOLLLEKPFQIFRAE 120  
 DB 61 ARGRIVSDQNMQILLSSIGDILLNKDLREELCGTEPCVLPFOLLLEKPFQIFRAE 120  
 QY 121 LMRVINDHSPYFLDREITLNLSTTPGATFLSADSDVGINNLNNTYISSNVYFHI 180  
 DB 121 LMRVINDHSPYFLDREITLNLSTTPGATFLSADSDVGINNLNNTYISSNVYFHI 180  
 QY 121 LMRVINDHSPYFLDREITLNLSTTPGATFLSADSDVGINNLNNTYISSNVYFHI 180  
 DB 121 LMRVINDHSPYFLDREITLNLSTTPGATFLSADSDVGINNLNNTYISSNVYFHI 180  
 QY 181 NVHDGEGNVSELYLDVLDREVEPELRLTLTGDDGSPRSRGLTIRIIVLINDNVP 240  
 DB 181 NVHDGEGNVSELYLDVLDREVEPELRLTLTGDDGSPRSRGLTIRIIVLINDNVP 240  
 QY 241 EFVESLYKVOYSPVSLVVTYSARDLDTGSGNGEIVYAFYATERTLKTFRINSTGN 300  
 DB 241 EFVESLYKVOYSPVSLVVTYSARDLDTGSGNGEIVYAFYATERTLKTFRINSTGN 300  
 QY 241 DEVBSLYKVOYSPVSLVVTYSARDLDTGSGNGEIVYAFYATERTLKTFRINSTGN 300  
 DB 241 DEVBSLYKVOYSPVSLVVTYSARDLDTGSGNGEIVYAFYATERTLKTFRINSTGN 300  
 QY 301 LHLKELNVEAIQYTLTLTQAKDGGSLGCKTVVHVTDINDPELIMSLTSPIDENS 360  
 DB 301 LHLKELNVEAIQYTLTLTQAKDGGSLGCKTVVHVTDINDPELIMSLTSPIDENS 360  
 QY 301 LHLKELNVEAIQYTLTLTQAKDGGSLGCKTVVHVTDINDPELIMSLTSPIDENS 360  
 DB 301 LHLKELNVEAIQYTLTLTQAKDGGSLGCKTVVHVTDINDPELIMSLTSPIDENS 360  
 QY 361 PETVAVERIRDRSGNNAKWCSTQDHLPEVLKPSVNEFTYTERALDDEERTENIT 420  
 DB 361 PETVAVERIRDRSGNNAKWCSTQDHLPEVLKPSVNEFTYTERALDDEERTENIT 420  
 QY 421 ITVVDLGPRLKTOHNLTVTVSDVNDNAPTESQTYTLVRRENSPFLHIGSVSATDRDS 480  
 DB 421 ITVVDLGPRLKTOHNLTVTVSDVNDNAPTESQTYTLVRRENSPFLHIGSVSATDRDS 480  
 QY 481 GANOVYSLPRLKTOHNLTVTVSDVNDNAPTESQTYTLVRRENSPFLHIGSVSATDRDS 540  
 DB 481 GANOVYSLPRLKTOHNLTVTVSDVNDNAPTESQTYTLVRRENSPFLHIGSVSATDRDS 540  
 QY 541 LSSQALVAVLADANDNAPFVLYPLONGSAPCTELVPAAGLIVAKVAVVDDSSQNA 600  
 DB 541 LSSQALVAVLADANDNAPFVLYPLONGSAPCTELVPAAGLIVAKVAVVDDSSQNA 600  
 QY 601 WLSTYLLKATPEGLFGVAHNGEVRTA 627  
 DB 601 WLSTYLLKATPEGLFGVAHNGEVRTA 627

RESULT 2

ID CDB2\_HUMAN STANDARD; PRT; 798 AA.  
 AC Q9Y5E7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protocadherin beta 2 precursor (PCDH-beta2).  
 GN PCDHB2.

OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99308636; PubMed=10380929;  
 RA Wu Q., Maniatis T.;  
 RT "A striking organization of a large family of human neural cadherin-  
 RL cell 97:779-790(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21223055; PubMed=11322959;  
 RA Vanhalst K., Kools P., Vanden Eynde E., van Roy F.;  
 RT "The human and murine protocadherin-beta one-exon gene families show  
 RT high evolutionary conservation, despite the difference in gene  
 RT number.";  
 RL PNAS Lett. 495:120-125(2001).  
 CC -1- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC -----  
 DR EMBL: AF152495; AAD43756.1; -;  
 DR EMBL: AF217756; AAK51623.1; -;  
 DR HSSP: P15116; INCU.  
 DR GeneW: HGNC:8687; PCDHB2.  
 DR MIM: 606328; -;  
 DR MIM: 604967; -;  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin; 5.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART: SM00112; CA; 6.  
 DR PROSITE: PS00232; CADHERIN\_1; 5.  
 DR PROSITE: PS00268; CADHERIN\_2; 6.  
 KW Calcium-binding; Cell adhesion; Glycoprotein; Signal; Repeat;  
 KW Transmembrane; Multigene family.  
 FT SIGNAL 1 30  
 FT CHAIN 1 798  
 FT DOMAIN 31 692  
 FT TRANSMEM 693 713  
 FT DOMAIN 714 798  
 FT DOMAIN 37 135  
 FT DOMAIN 136 244  
 FT DOMAIN 249 349  
 FT DOMAIN 354 453  
 FT DOMAIN 458 563  
 FT DOMAIN 570 673  
 FT CARBOHYD 171 171  
 FT CARBOHYD 420 420  
 FT CARBOHYD 438 438  
 FT CARBOHYD 569 569  
 SQ SEQUENCE 798 AA: 87253 MW: 82602865378CB27 CRC64;  
 Query Match 72.4%; Score 2323; DB 1; Length 798;  
 Best Local Similarity 73.0%; Pred. No. 1.7e-14;  
 Matches 451; Conservative 65; Mismatches 102; Indels 0; Gaps 0;





QY 481 GANAQVTSLLPDPHPLGLSIVSINADNGQLFALRSIDFEALQAEFFRYGAADRGSPA 540  
 DB 481 GNAQVNTSLPDPHPLGLSIVSINADNGQLFALRSIDFEALQAEFFRYGAADRGSPA 540  
 QY 541 LSSQALVRYLVADANDANPFLVYPLQNGSAPCTELVPRAPAEGLYVAVAVDGDGQNA 600  
 DB 541 LSSEALVRYLVADANDANPFLVYPLQNGSAPCTELVPRAPAEGLYVAVAVDGDGQNA 600  
 QY 601 WLSTQLKATPEGLFGVMAHNGEVRTA 627  
 DB 601 WLSTQLKATPEGLFGVMAHNGEVRTA 627

RESULT 4  
 CDBF\_HUMAN STANDARD: PRT: 787 AA.  
 AC Q9Y5E8: 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protocadherin beta 15 precursor (PCDH-beta15).  
 GN PCDHB15.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99308636; PubMed=10380929;  
 RA Wu Q., Maniatis T.;  
 RT "A striking organization of a large family of human neural cadherin-  
 RL like cell adhesion genes";  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21223055; PubMed=11322959;  
 RA Vanhaest K., Kools P., Vanden Eynde E., van Roy F.;  
 RT "The human and murine protocadherin-beta one-exon gene families show  
 RT high evolutionary conservation, despite the difference in gene  
 RL number";  
 [2]  
 RL FEBS Lett. 495:120-125(2001).  
 CC -1- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC EMBL: AF152494; AAD43755.1;  
 DR EMBL: AF217742; AAK51610.1;  
 DR Genew: HGNC:8686; PCDHB15.  
 DR MIM: 606341;  
 DR MIM: 604967;  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin.5.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART: SM00112; CA.5.  
 DR PROSITE: PS00232; CADHERIN.1; 5.  
 DR PROSITE: PS0268; CADHERIN.2; 6.  
 KW Calcium-binding; Cell adhesion; Glycoprotein; Signal; Repeat;  
 KW Transmembrane; Multigene family.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 787 PROTOCADHERIN BETA 15.  
 FT DOMAIN 27 690 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 691 711 POTENTIAL.  
 FT DOMAIN 712 787 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 713 787 CADHERIN 1.

FT DOMAIN 138 242 CADHERIN 2.  
 FT DOMAIN 247 347 CADHERIN 3.  
 FT DOMAIN 352 451 CADHERIN 4.  
 FT DOMAIN 456 561 CADHERIN 5.  
 FT DOMAIN 568 671 CADHERIN 6.  
 FT CARBOHYD 418 418 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 567 567 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 SQ SEQUENCE 787 AA; 86328 MW; 8DB8D3A07B0BA56 CRC64.

Query Match 71.3%; Score 2287; DB 1; Length 787;  
 Best Local Similarity 72.3%; Pred. No. 4, 2e-145;  
 Matches 447; Conservative 67; Mismatches 104; Indels 0; Gaps 0;

QY 10 QKROVSLCVFLGVNAGAEPLRYFAEETERTGLANLAIDLGLGVEELASRGCRIVSD 69  
 DB 10 EQQVILLILLLEVLTLAGEPRRYSMEEFEGSFYANLANDGLGVGLARGARVASE 69  
 QY 70 ETIGFLINPLTDLLNKLREELCGTPECVLPFOILLKPKPOIPFAELVAVDINDH 129  
 DB 70 DNEQGLQDLQTLQGLLNLKREKLCGPTPEIMHFVLLKPLEVFRAELLVTDINDH 129  
 QY 130 SPVFLDREITLNLSTFTGAPFLAESADSDVGINNLNNTYSSNVYHINVHNGEGN 189  
 DB 130 SPFFPEREMTKIPETSSLSGLTVPLKARDLVGSMNNGYINISPSHHVSTRRGDSR 189  
 QY 190 VSEFLVDKYLDEEVEPELRLTLGLDGSPPRSSTTLRLVLVDINDVPEVESLYK 249  
 DB 190 KPELVLDPELDREEOAEELRLTLAVDGSPPRSSTVQLLLVLANDNAPEFVQALYEV 249  
 QY 250 QVENSPPVGLVTVYARLDLDSNGEITYAFYATERTLKTFRINSTGNHLKAEIN 309  
 DB 250 QVENSPPVGLVTVYARLDLDSNGEITYAFYATERTLKTFRINSTGNHLKAEIN 309  
 QY 310 EIATQTLTFOAKDGGGLSGKCTVVVHTVDINDNPBELLMSLSPSPENSPETVAVER 369  
 DB 310 ETMSVTDLDLEASDGGGLSGKCSVSKYLVVDNPFELSLSPSPENSPETVALFR 369  
 QY 370 IRDRSGNNAKVCSIQDHLPEVLKPSVENEFTLVTERALDRBERTVNTTIVDGLTP 429  
 DB 370 IRDRSGNNAKVCSIQDHLPEVLKPSVENEFTLVTERALDRBERTVNTTIVDGLTP 429  
 QY 430 RLKQHNLTVTSDVDNAPTEPOTTYTLRVRRNNSPALHISVSATDSDGANNQVYS 489  
 DB 430 RLKQHNLTVTSDVDNAPTEPOTTYTLRVRRNNSPALHISVSATDSDGANNQVYS 489  
 QY 490 ILPPHDPOLPLGSLVSIADNGQLFALRSIDFEALQAEFFRYGAADRGSPALSVAV 549  
 DB 490 ILPPHDPOLPLGSLVSIADNGQLFALRSIDFEALQAEFFRYGAADRGSPALSVAV 549  
 QY 550 LVADANDANPFLVYPLQNGSAPCTELVPRAPAEGLYVAVAVDGDGQNAVLSTQLKA 609  
 DB 550 LVADANDANPFLVYPLQNGSAPCTELVPRAPAEGLYVAVAVDGDGQNAVLSTQLKA 609  
 QY 610 TERGLFGVMAHNGEVRTA 627  
 DB 610 TERGLFGVMAHNGEVRTA 627

RESULT 5  
 CDB3\_HUMAN STANDARD: PRT: 796 AA.  
 AC Q9Y5E6: 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protocadherin beta 3 precursor (PCDH-beta3).  
 GN PCDHB3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RP SEQUENCE FROM N.A.

RX MEDLINE-9308636; PubMed-10380929;  
 RA Wu O., Maniatis T.;  
 RT "A striking organization of a large family of human neural cadherin-  
 RT like cell adhesion genes.";  
 RL Cell 97:779-790(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-21223055; PubMed-11322959;  
 RA Vanhalst K., Kools P., Vanden Eynde E., van Roy F.;  
 RT "The human and murine protocadherin-beta one-exon gene families show  
 RT high evolutionary conservation, despite the difference in gene  
 RT number.";  
 RL FEBS Lett. 495:120-125(2001).  
 CC -1- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AF152496; AAD43757.1; -  
 CC EMBL: AF217755; AAK51622.1; -  
 CC Genew: HGNC:8688; PCDBB3.  
 CC MIM: 606329; -  
 CC MIM: 604967; -  
 CC InterPro: IPR002126; Cadherin.  
 CC Pfam: PF00028; cadherin; 5.  
 CC PRINTS: PRO0205; CADHERIN.  
 CC SMART: SM00112; CA; 6.  
 CC PROSITE: PS00232; CADHERIN\_1; 5.  
 CC PROSITE: PS00268; CADHERIN\_2; 6.  
 CC DR Calcium-binding; Cell adhesion; Glycoprotein; Signal; Repeat;  
 CC KW Transmembrane; Multigene family.  
 CC FT SIGNAL 1 26  
 CC FT CHAIN 27 796  
 CC FT DOMAIN 27 690  
 CC FT TRANSMEM 691 711  
 CC FT DOMAIN 712 796  
 CC FT DOMAIN 133 133  
 CC FT DOMAIN 138 242  
 CC FT DOMAIN 247 347  
 CC FT DOMAIN 352 451  
 CC FT DOMAIN 456 561  
 CC FT DOMAIN 568 671  
 CC FT CARBOHYD 169 169  
 CC FT CARBOHYD 418 418  
 CC FT CARBOHYD 436 436  
 CC FT CARBOHYD 567 567  
 CC SO SEQUENCE 796 AA; 86772 MW; F586CF1BC3A2B0D2 CRC64;  
 Query Match 70.5%; Score 2262; DB 1; Length 796;  
 Best Local Similarity 71.3%; Pred. No. 2e-143;  
 Matches 447; Conservative 63; Mismatches 117; Incls 0; Gaps 0;  
 OY 1 MEARVVAHAKQKRVSLCYFLGVSNAGAPRLRYFAEERGCFFLANLAIDGLGVEELS 60  
 DB 1 MEGGGRFLKROVLLFFVLGSLAGSRRYSVAEEKKFLANLAKDLGLRVEELA 60  
 OY 61 ARGRIVSDETIGFLINPLDGLINLEKIDREELCGPEPCVLPDPOLLLEKPFIFRAE 120  
 DB 61 ARGAQVYSGNKHQPLSHQTDGLINLEKIDREELCGPEPCVLPDPOLLLEKPFIFRAE 120  
 OY 121 LWRINDHSPVFLDREITLNLLESTPGATFLLESADSDVGINNLKNTYITSSNVYEH 180  
 DB 121 LRIIDNDHSPVFEENEMLKILESTLPCTVPLGNAEDLDVGRNSLQWYITTPSHFV 180  
 OY 181 NVHDNCEGVNSELVDKYLDEEVEPELTLTGLDGSGPPSSGTTLLRIVLVDINDNP 240

DB 181 LFRSRDRGKRPVELYLDALDREPELSTLTALDGGSPRRSGTAQINQIYDINDNP 240  
 OY 241 EEVESLKYQVEENSPVSLVTVTSARDLDGSGNEIYAFYATERTLKTFRRINSTGN 300  
 DB 241 EFAQPLYEVALENTPVNSVITVASDLDGSGFEGTISYAFPHASEEIRKTFOLNPIGD 300  
 OY 301 LHLKELNVEATQIOTYTLTLOAKDGGGLSGKCVVYVHVNDINPPELLMSSLTSPSPNS 360  
 DB 301 MOLVYINFEAINSEVDIEAKDGGGLSGKSTVYQVVDVNDNPPELLSVNSPSPNS 360  
 OY 361 PETVAVERIRDRDGNNAKVCSTQDHLPEVLPKPSVNEFTLVYERALDREERTENIT 420  
 DB 361 GGTVALVESVDLDGSGDGRVWCSTIENMLPEFKPSVNEFTLVSEGLDRETSSEVIT 420  
 OY 421 ITVTDLGPRKLTQNNLTVTSVDVNDNAPTSTQTYTTRVKNNSPALHTSVSATDSDS 480  
 DB 421 ITITDLGPRKLTQNNLTVTSVDVNDNAPTSTQTYTTRVKNNSPALHTSVSATDSDS 480  
 OY 481 GANAQVTSILPPIHPOLPLGSLVSIINDNGOLFALSLDEPLAQAPEFRGADRGSPA 540  
 DB 481 GINAQVTSILPPIHPOLPLGSLVSIINDNGOLFALSLDEPLAQAPEFRGADRGSPA 540  
 OY 541 ISSQALVRVLVADANDNAPFVLYPLONGSAPCTELVPRAAEAGYLVAKVAVDGDGCGNA 600  
 DB 541 ISSEALVRVLVADANDNAPFVLYPLONGSAPCTELVPRAAEAGYLVAKVAVDGDGCGNA 600  
 OY 601 WLSTYQLKATEPGLGVAHNGEVRTA 627  
 DB 601 WLSTYQLKATEPGLGVAHNGEVRTA 627  
 RESULT 6  
 CDBG\_HUMAN STANDARD; PRT; 776 AA.  
 AC Q9NKCJ7; G9HCE1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protocadherin beta 16 precursor (PCDH-beta16) (Protocadherin 3x).  
 GN PCDBB16 OR PCDH3X OR KIAA1621.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kools P.F., J., Van Roy F.;  
 RT "The human protocadherin 3 gene cluster, expression analysis of these  
 RT one-exon genes";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE-20450683; PubMed-10997877;  
 RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes.  
 RT XVIII. The complete sequences of 100 new cDNA clones from brain which  
 RT code for large proteins in vitro.";  
 RL DNA Res. 7:273-281(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Kools P.F., J., van Roy F.M.;  
 RT "Molecular analysis of the human protocadherin-3 (PCDH-Beta) gene  
 RT cluster";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF217757; AAF81914.1; -  
 DR EMBL: AB046841; BAB13447.1; ALT\_INIT.  
 DR EMBL: AF282973; AAC10030.1; -  
 DR Genew; HGNC:14546; PCDBH16.  
 DR MIM: 606345; -  
 DR MIM: 604967; -  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin.5.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART; SM00112; CA: 5.  
 DR PROSITE; PS00232; CADHERIN\_1; 5.  
 DR PROSITE; PS0268; CADHERIN\_2; 6.  
 DR Calcium-binding; Cell adhesion; Glycoprotein; Signal; Repeat;  
 KW Transmembrane; Multigene family.  
 FT SIGNAL 1 28  
 FT CHAIN 29 776  
 FT DOMAIN 29 690  
 FT TRANSMEM 691 711  
 FT DOMAIN 712 776  
 FT DOMAIN 35 133  
 FT DOMAIN 138 242  
 FT DOMAIN 242 347  
 FT DOMAIN 347 451  
 FT DOMAIN 451 561  
 FT DOMAIN 561 671  
 FT CARBOHYD 418 418  
 FT CARBOHYD 436 436  
 FT CARBOHYD 567 567  
 FT CONFLICT 482 482  
 FT CONFLICT 525 526  
 FT CONFLICT 532 532  
 FT CONFLICT 543 543  
 FT CONFLICT 638 638  
 FT CONFLICT 652 652  
 FT CONFLICT 710 710  
 FT SEQUENCE 776 AA; 84935 MW; 93EF1C641A1DE0FB CRC64;  
 Query Match 70.1%; Score 2250; DB 1; Length 776;  
 Best Local Similarity 69.5%; Pred. No. 1,2e-142;  
 Matches 436; Conservative 74; Mismatches 117; Indels 0; Gaps 0;

QY 1 MEARVHALQKROVSLICVFLGVSWAGAPLRVFAETERTGTEFLANLADLGLVEELS 60  
 DB 1 MEIGMHNRROVAVFVLLSLSGAGALGSAVETERTGSEFVANGKDLGLTEMS 60  
 QY 61 ARGCRVSDERITGLLNPGLDILLNKLDRRELCTGTEPCVLPQOLLLEKPOIFRAE 120  
 DB 61 TRKATISQGNKQHLQLAQGGDLLINKEDRELCTGTEPCILHFOVLMENPLEIQAE 120  
 QY 121 LWNVDINDHSHVFLDREITLILSTTPGATFLLESADHSDVGINLRNTYISSNYFHI 180  
 DB 121 LRAVIDINDHSMFTEKEKILKIPENSPGTEFPLNHALDDVGSNNQNTYISSSHFRV 180  
 QY 161 NVHDNGEENYSELVLDVLDREVEPELRLLTGLDGGSPRSSTGLIRILVIDINDNP 240  
 DB 161 LIHERDRKRYPELVLDELREPEPQLRLLTALDGGSPRSSTAQRILEVVIDINAP 240  
 QY 241 EVESTLYVOYVSPVSLVVTVSARDLDSNGELIYAFYFATERTKTRFNINSNGN 300  
 DB 241 EEOPTIYVQIPENSPPLSLVATVSARDLDSGANGKISYLFQSEDISKLEVENPMTE 300  
 QY 301 LHLKALNVEAIIQYTLTIOAKDGGSLGKCTVVVHVVDINDNPELMSLSPIPENS 360  
 DB 301 VRLRKQVDFEAVTSEVRIKATDGGSLGKCTLLQVVDVNDNPPQYTMALSPIPENS 360  
 QY 361 PETVAVVRIRDRSGNNAKVCISIQHLPFLVKPSYENFVTLVTERALDREERTENIT 420  
 DB 361 PEIYVAVSVSDPDGSGNGKTISSIQEDLPFLKPSVKNFYTLVTERALDREAREENIT 420

QY 421 ITVTDLGTEPLKTOHNLTVTVSDVNDNAPTFQSTYTLFRVNNSPALHIGSVSATDRDS 480  
 DB 421 LVTVDGTEPLKTEHNITVOISDVNDNAPLFTQSTYTLFRVNNSPALHIGSVSATDRDS 480  
 QY 481 GNAQVYTSILPPHPDQLPLGSLVSTNANGQLFAIRSIDFALQAFERRVCAARGSPA 540  
 DB 481 GTNAQVYTSILPPQDHPPLASLVSINADNGHIFALRSIDYBALQAFERRVCAARGSPA 540  
 QY 541 LSSQALVRLVADANDNAPFVLYPLONGSAPCTELVPRAAGVYAVKVAVDGSDGONA 600  
 DB 541 LSRREALVRLVADANDNSFVLYPLONGSAPCTELVPRAAEFGYLVTKVAVDGSDGONA 600  
 QY 601 WLSYQLKATEPGLFGVMAHNGEVRTA 627  
 DB 601 WLSYQLKATEPGLFGVMAHNGEVRTA 627

## RESULT 7

CDBA\_HUMAN STANDARD; PRT; 800 AA.  
 ID CDBA\_HUMAN  
 AC Q9UN67;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protocadherin beta 10 precursor (PCDH-betac10).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxId:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-99308636; PubMed-10380929;  
 RX Wu Q., Maniatis T.;  
 RT "A striking organization of a large family of human neural cadherin-  
 like cell adhesion genes."  
 RL Cell 97:779-790(1999).  
 CC -1- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF152489; AAD43750.1; -  
 DR Genew; HGNC:8681; PCDBH10.  
 DR MIM: 606336; -  
 DR MIM: 604967; -  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin.5.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART; SM00112; CA: 5.  
 DR PROSITE; PS00232; CADHERIN\_1; 5.  
 DR PROSITE; PS0268; CADHERIN\_2; 6.  
 DR Calcium-binding; Cell adhesion; Glycoprotein; Signal; Repeat;  
 KW Transmembrane; Multigene family.  
 FT SIGNAL 1 36  
 FT CHAIN 27 800  
 FT DOMAIN 27 692  
 FT TRANSMEM 693 713  
 FT DOMAIN 714 800  
 FT DOMAIN 35 133  
 FT DOMAIN 138 242  
 FT DOMAIN 242 347  
 FT DOMAIN 347 451  
 FT DOMAIN 451 561  
 FT DOMAIN 561 671  
 FT CARBOHYD 169 169  
 FT SEQUENCE 800 AA; 84935 MW; 93EF1C641A1DE0FB CRC64;  
 Query Match 70.1%; Score 2250; DB 1; Length 776;  
 Best Local Similarity 69.5%; Pred. No. 1,2e-142;  
 Matches 436; Conservative 74; Mismatches 117; Indels 0; Gaps 0;



```

Db 61 ARCTRYVSDNKKYLLLDSTGTLNKLKREKLCGPEPCMTLFTQILMDPFOIYRAE 120
QY 121 LWRVDINDSPVFLDREITLNLIESTTPGATFLLSASHDSDVGINLRYTISNVEFRI 180
Db 121 LVRVDINDSPVFRHKEMLKISENTAEGTAFLEAOPDDEGSHNSIQWYTISSNFRI 180
QY 181 NVHDNGEGSESLVDKVLDRREVEBELRLTLTGDSGSPRSCTTLRIIVLDINDNP 240
Db 181 KISGSEGMATYPELVLDKDLREDEBELSLTLALDGGSPSSGSTTRIVVLDVNDAP 240
QY 241 EFVESLYKQVPEPNSVGLVYVVSARDLDTGSNGEIVAFYATERTKTPRINSTGN 300
Db 241 QFQALYETQAPENSPVGLIYKVSAGDAGCVNAEVSSEFDABEDILTTQINPFGE 300
QY 301 LHKALNTEAQTTLTQADGGGLSKCTVYVHVHTDINDNPPELLMSSITSPENS 360
Db 301 IFRRELDYELVNSYKINQADGGGLSARCTVLKVLDSNPNPELLISSLNSVAENS 360
QY 361 PETVAVFRIRDRSGNNAKVCSIODHLPEVLKPSVENFYTLVTERALDREERTYNT 420
Db 361 PGVLAVERKIKDRDSEKNTCTCYODNLPFLKPSVDNFYTLMEGALDRSKAENYNT 420
QY 421 IYVTDLGPRLKTQNLVTVSDVNDNAFTFSQTTYYTLRVRENSPALHIGSVATDRS 480
Db 421 IYVTDLGPRLKTEHSTLQVSDVNDNAPAFQTSYTLFVRENSPALHIGSVATDRS 480
QY 481 GNAQVTVSLRPHDPQLPLGSLVSNADNGOLFALRSIDFALOAFTERGADRGSPA 540
Db 481 GNAQVTVSLRPHDPQLPLGSLVSNADNGOLFALRSIDFALOAFTERGADRGSPA 540
QY 541 LSSQALVRLVADANDNAPEVLYPLONGSAPCTELVPRAAEGYVAVVAVDGGGNA 600
Db 541 LSSQALVRLVADANDNAPEVLYPLONGSAPCTELVPRAAEGYVAVVAVDGGGNA 600
QY 601 WLSYOLKATKEPGLFGVAHNGEVRTA 627
Db 601 WLSYOLKATKEPGLFGVAHNGEVRTA 627

RESULT 9
CDB4_HUMAN STANDARD; PRT; 795 AA.
ID CDB4_HUMAN
AC Q9Y5E5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protocadherin beta 4 precursor (PCDH-beta4).
GN PCDH4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9308636; PubMed=10380929;
RA Wu Q., Maniatis T.;
RT "A striking organization of a large family of human neural cadherin-
RL like cell adhesion genes.";
RT Cell 97:779-790(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=2122305; PubMed=11322959;
RA Vanhaest K., Kools P., Vanden Eynde E., van Roy F.;
RT "The human and murine protocadherin beta one-exon gene families show
RT high evolutionary conservation, despite the difference in gene
RT number.";
RL FEBS Lett. 495:120-125(2001).
CC -1- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (by similarity).
CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF152497; AAD43758.1; -
CC EMBL; AF217754; AAK51621.1; -
CC HSSP; P15116; INCU.
CC Genew; HGNC:8689; PCDH4.
CC MIM; 606330; -
CC MIM; 604967; -
CC InterPro; IPR002126; Cadherin.
CC Pfam; PF0028; cadherin; 5.
CC PRINTS; PR00205; CADHERIN.
CC SMART; SM00112; CA; 5.
CC PROSITE; PS00232; CADHERIN_1; 5.
CC PROSITE; PS00268; CADHERIN_2; 5.
CC Calcium-binding; Cell adhesion; Glycoprotein; Signal; Repeat;
CC Transmembrane; Multigene family.
CC SIGNAL 1 27
CC CHAIN 28 795
CC DOMAIN 28 689
CC TRANSMEM 690 710
CC DOMAIN 711 795
CC DOMAIN 34 132
CC DOMAIN 137 241
CC DOMAIN 246 346
CC DOMAIN 351 450
CC DOMAIN 455 560
CC DOMAIN 567 670
CC CARBOHYD 183 183
CC CARBOHYD 417 417
CC CARBOHYD 435 435
CC CARBOHYD 566 566
CC SEQUENCE 795 AA; 87269 MW; 32A521247DD6A56F CRC64;

Query Match 68.7%; Score 2205; DB 1; Length 795;
Best Local Similarity 69.5%; Pred. No. 1,3e-139;
Matches 428; Conservative 70; Mismatches 118; Indels 0; Gaps 0;

QY 12 ROYSLICVFLGVMAGAEPLRYAFVARETERGFLAMLAIDIGVLELSARCRIVSDPT 71
Db 11 ROYLAFLMFLVLSQVRLERISVLETSVGAHAAOLGIGELASRARVLSDD 70
QY 72 IGFLLNPLTGDLINLNEKLDREELCGPTEPCVFPOLLKEFQIIRAEIWRVINDISP 131
Db 71 KORLODRQTDGLLREKLDREELCGPIPCVHGFVLEMPVQFQSGILLQDINDISP 130
QY 132 VFIDREITLNLIESTTPGATFLLSASHDSDVGINLRYTISNVEFRIHNDNGENVY 191
Db 131 IPERREVLKLIENSOPGLFPLLIADDDVGSNGLOKTYTISPNSHSEKKY 190
QY 192 SELVLDKVLDRREVEBELRLTLTGDSGSPRSCTTLRIIVLDINDNVEFESLYKQV 251
Db 191 PDLVQKPLDREQPEFSLVLAALDGGSPRSKGTWVRLINDINDNAEFVHTPYGVQV 250
QY 252 PENSVPGLVTVVSARDLDTGSNGEIVAFYATERTKTPRINSTGNHLKALNEA 311
Db 251 LENSPLDSPIVRLARDIDAGNFGSVYGLFQASDEIKOTFESINEVTEGILLKKLDEK 310
QY 312 IQTYTLTQADGGGLSGCTVAVVHVHTDINDNPPELLMSSITSPENSPETVAVFRIR 371
Db 311 IKSYPEIATDGGGLSGKGTAVIEVDVNDNPPELLISSLSPENAPETVAVFRIR 370
QY 372 DRDGGNNAKVCSIODHLPEVLKPSVENFYTLVTERALDREERTYNTITVTDGTRRL 431
Db 371 DRDGGNGMKISIDNPFLIKPKTEYTLVTRPDRDSAEYNTITVTDGTRRL 430
QY 432 KTOHNLTVSVVNDNAPFTFSQTTYYTLRVRENSPALHIGSVASDRDGGGNAQVTVSL 491
Db 431 KTOQNTTVSVVNDNAPFTFSQTTYYTLRVRENSPALHIGSVASDRDGGGNAQVTVSL 490

```

[illegible]







Query Match	67.88;	Score 2174.5;	DB 1;	Length 794;
Best Local Similarity	68.68;	Pred. No. 1.4e-137;		
Matches 430; Conservative	69;	Mismatches 125;	Indels 3;	Gaps 2

QY 1 MEAVVHAALOKROVSLICVFLOGVSNAGAEPLRYEVAETTERGTEFLANLAIDLIGVEELS 60  
 :  
 Db 2 MOTKVOV--KKRVOAFEIIILMGEVGSESIQYVLAEETSGTFVANIKDLGRNGELA 59  
 :  
 QY 61 ARGRIVSDTISGFLLNPLDGLLNLKEDRLKEELCGTEPCVAPFOULLEKPPQLFRAE 120  
 :  
 Db 60 SRGRVVEFGKNROHLQFDPOTHDLLNKERLEELCGSTECVCAPFOVLEENP,QEPQOAS 119  
 :  
 QY 121 LKWADINDHSPVLDREITLNTILESTTPGAATFLSAEDSVGINNI,RNTISSNVYFH 180  
 | :  
 Db 120 LRARDINDAHPPEPAEMLIKISELTMPGKITFLPKAMANDLDTGSNGCORTISSNPFFHV 179  
 :  
 QY 181 NVHDNEGNYVELVDKYLVDREEVPBELRTLTLTGIDGSGSPRGSTFLIRLVLDINDNPV 240  
 :  
 Db 180 LTRRSBGRKFPELVLDKPLDREQPOLRTLIALDGSSPRGSTSELQYOVLINDNPV 239  
 :  
 QY 241 EPEBSLKYYOVPNBSVGSVLVTVSARDDLDTGSNGEIYAEFYATERTLKTFRINSTGN 300  
 | :  
 Db 240 EFQAOLTEYAQVPENNPIGSLVTIVTSABDLDGSEFKGYALFOYDD--VNOPFEINATIGE 298  
 :  
 QY 301 LHKAELNVALIOTYTTLTIOAKDGGISGSKCYVVYHVATDINDNPPELLMSLSIPENS 360  
 :  
 Db 299 IRLKALDAFEIIDSUYDVDAIDGGISGSKCSLVYVRVLDVNDAPLELTJMSFFSLIPENI 358

RESULT 13		
ID	CDBB_HUMAN	STANDARD; PRT; 797 AA.
AC	Q9Y5F2;	
DT	16-OCF-2001 (Rel. 40, Created)	
DT	16-OCF-2001 (Rel. 40, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Protocadherin beta 11 precursor (PCDH-beta11).	

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99308636; PubMed=10380929;
RA Wu Q., Maniatis T.;
RT "A striking organization of a large family of human neural cadherin-
RT like cell adhesion genes.";
RL Cell 97:779-790(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21223055; PubMed=11329599;
RA Vanhalst K., Koels P., Vanden Eynde E., van Roy F.;
RT "The human and murine protocadherin beta one-exon gene families show
RT high evolutionary conservation, despite the difference in gene
RT number.";
RL FEBS Lett. 495:120-125(2001).
CC -1- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF152490; AAD33751.1; -
DR EMBL; AF217747; AAK51615.1; -
DR HSSP; P15116; 1NCJ
DR Genew; HGNC:8682; PCDHB11.
DR MIM; 606337; -
DR MIM; 604967; -
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 6.
DR PROSITE; PS00232; CADHERIN_1; 5.
DR PROSITE; PS50268; CADHERIN_2; 6.
DR

```

KW Calcium-binding; Cell adhesion; Glycoprotein; Signal; Repeat;  
 Transmembrane; Multigene family.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 797 PROTOCADHERIN BETA 11.  
 FT DOMAIN 27 690 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 691 711 POTENTIAL.  
 FT DOMAIN 712 797 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 35 133 CADHERIN 1.  
 FT DOMAIN 138 242 CADHERIN 2.  
 FT DOMAIN 247 347 CADHERIN 3.  
 FT DOMAIN 352 451 CADHERIN 4.  
 FT DOMAIN 456 561 CADHERIN 5.  
 FT DOMAIN 568 671 CADHERIN 6.  
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 436 436 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 487 487 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 797 AA; 87087 MW; 8FBA0DBC2858504 CRC64;

Query Match 67.7%; Score 2171; DB 1; Length 797;  
 Best Local Similarity 68.3%; Pred. No. 2,3e-137;  
 Matches 428; Conservative 76; Mismatches 123; Indels 0; Gaps 0;

QY 1 MEARYVHALOKROYSLCVPFLGVSNAGAPLRYFAEETEGTFLANALDGLCEELS 60  
 DB 1 MENOQTRQOIRQVLLFLVLMGMSQAGEIWSFSVAEEMOSGFVGNLAKDKLVRELS 60  
 QY 61 ARGGRVSDETIGFLNPLTGLDILLNKLREELCGTEPCVLPFQOLLEKQIFRAE 120  
 DB 61 SRGARVSNDRKKQRLQDLDINTGDLISLSELDREELCGSEPCVHLQVLMQNPQFLQIE 120  
 QY 121 LMRVDINDHSVFTDRETTNILESTTGATFLLESASDVGINNLNRYTSSNVIYHI 180  
 DB 121 LQVRINDHSIFSEKQMLTEIPENSIVGAVFLLESASDVLGYINAVKSYITSPSHFHI 180  
 QY 181 NVHNGEGNVSELYLVDRREYELRLTLTGDSGSPRSCTTLRIITLVDINDVNP 240  
 DB 181 KMRVYIPDRKKPELVLDALDYELPELSTLSALDGSPPRSCTALRVVVDINDNSP 240  
 QY 241 EFVESLKYQVQENSPVSLVTVVSARDLGTSGEIVYAFYATERTLKTFRINSTSGN 300  
 DB 241 EFEQAFYEVKIRENSILIGSLIIVSAMOLDGTNGELCYTFSHMSDRLKRFETINQKGE 300  
 QY 301 LHKAEALVEAIQYTLTIOAKDGGGLSGKCTVYVHTVDINDNPELIMSSITSPENS 360  
 DB 301 ITRAPLDEFETIESYIIQATDGGGLGKSTVIIVHDVNDNAPELIVSSITSPENIT 360  
 QY 361 PETVAVAFRIDRDSGNNAKWCSTIODLPEVLKPSVENETVYTERALDSEEREYVIT 420  
 DB 361 PETVAVAFSIODIDSGNGRIVCSIPEDLPVYLKSSVENITTLTERFLDNESTAEVIT 420  
 QY 421 ITVVDLGTPLKTOHNLTVVSDVNDNAPTESQTTYTLRVENNSPALHIGSVSATDRDS 480  
 DB 421 ITVVDLGTPLKTEHNTVVLVSDVNDNAPTFOTISYTLFVRENNSPALHIGSVSATDRDS 480  
 QY 481 GANNOVYTSILPHDPOLPLGSLVSYINDNGQLPALRSLDEALQAEFRGADRGSPA 540  
 DB 481 GTNAOVNYSILPDLHLPLASLVSINFDNGHLPALRSLDEALQAEFRGADRGSPA 540  
 QY 541 LSSQALVAVLADANDNAPFLVLPKONGSAPCTELVPRAAGYVAVVAVDGGSGNA 600  
 DB 541 LSSQALVAVLADANDNAPFLVLPKONGSAPCTELVPRAAGYVAVVAVDGGSGNA 600  
 QY 601 WLSTYOLKATEPGLFGVAHNGEVRTA 627  
 DB 601 WLSTYOLKATEPGLFGVAHNGEVRTA 627  
 RESULT 14  
 CDB8\_HUMAN STANDARD; PRT; 801 AA.  
 ID CDB8\_HUMAN  
 AC Q9UN66;  
 DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN Protocadherin beta 8 precursor (PCDH-beta8) (Protocadherin 31).  
 OS PCDBH8 OR PCDH31.  
 OC Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99308636; PubMed=10380929;  
 RA Wu Q., Maniatis T.;  
 RT "A striking organization of a large family of human neural  
 RT cadherin-like cell adhesion genes.";  
 RN Cell 97:779-790(1999).  
 RP REVISIONS.  
 RA Wu Q., Maniatis T.;  
 RN Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 192-801 FROM N.A.  
 RA Kools P.F.J., van Roy F.M.;  
 RT "Molecular analysis of the human protocadherin-3 (PCDH-beta) gene  
 RT cluster.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF152501; AA043762.2; -  
 DR EMBL; AF282973; AA010031.1; -  
 DR Genew; HGNC:8693; PCDBH8.  
 DR MIM; 606334; -  
 DR MIM; 604967; -  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam; PF00028; cadherin. 5.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 6.  
 DR PROSITE; PS00232; CADHERIN\_1; 5.  
 DR PROSITE; PS0268; CADHERIN\_2; 5.  
 KW Calcium-binding; Cell adhesion; Glycoprotein; Signal; Repeat;  
 KW Transmembrane; Multigene family.  
 FT SIGNAL 1 28 POTENTIAL.  
 FT CHAIN 29 801 PROTOCADHERIN BETA 8.  
 FT DOMAIN 29 691 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 692 710 POTENTIAL.  
 FT DOMAIN 711 801 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 36 134 CADHERIN 1.  
 FT DOMAIN 139 243 CADHERIN 2.  
 FT DOMAIN 248 348 CADHERIN 3.  
 FT DOMAIN 353 452 CADHERIN 4.  
 FT DOMAIN 457 562 CADHERIN 5.  
 FT DOMAIN 569 672 CADHERIN 6.  
 FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 199 199 V -> L (IN REF. 3).  
 FT CONFLICT 234 234 V -> L (IN REF. 3).  
 FT CONFLICT 244 244 Q -> E (IN REF. 3).  
 FT CONFLICT 569 569 G -> S (IN REF. 3).  
 FT CONFLICT 639 639 Q -> H (IN REF. 3).  
 FT CONFLICT 661 661 L -> V (IN REF. 3).  
 FT CONFLICT 767 767 F -> L (IN REF. 3).  
 SO SEQUENCE 801 AA; 87623 MW; 30FB615CADA3A051 CRC64;

Query Match 67.2%; Score 2156.5; DB 1; Length 801;  
 Best Local Similarity 67.5%; Pred. No. 2.2e-136;  
 Matches 424; Conservative 68; Mismatches 135; Indels 1; Gaps 1;

QY 1 MEARVVALOKROVSLVFLGVSMAG-AEPLKRYFAEETERTFLANLAIDGLGVEEL 59  
 DB 1 MEASGLICROROVLSFLLGLSLAGAAEPKRSYVEETEGSSFTVNLADLGLEOREF 60

QY 60 SARCRIVSDETIGFLLNPLTGLDLNKLDEEELGPEPCVLPFOLLKPEFQIFRA 119  
 DB 61 SRGRVAVSRGKHLQNLQETADLLNEKLDKEDLCGHTPCVLRVYLLSEPEFQIA 120

QY 120 ELWVDINDHSPVFLDREITLNIESTTPGATFLLSAHSDVGINLNRNYTSSNVYFH 179  
 DB 121 ELQVIDINDHSPVFLDKQMLVKSSESPGTAFLKNAEDDIGNNIENIISNSYFR 180

QY 180 INVHNDGEGNVSELYVDKVDREVPRLTLTGDLGGSPRSGTTLIRLVLDINDV 239  
 DB 181 VLTFRKSDGRKPELVLDKLDREBEALRLTLALDGGSPRSGTAQVYIEVDVNDNA 240

QY 240 PEFEVSLKYQVENSPPVGLVTVSARDLDTGSGNEIYAFVAFERTLKTFRINSNG 299  
 DB 241 PEFGQPFYRVQISEDSPISEFLVYKSAIDVDVTGNGEISTSLFQASDEISTKFKVDLTG 300

QY 300 NLHLKAEINTEAIOYTLTLTQAKDGGSLGKCTVVVHTDINDPELLMSLTSPIEN 359  
 DB 301 EIRLKQDLEKQSYEVNIEARDAAGSGKCTVLIQYIDVNDHAFVETMAFSPIDEN 360

QY 360 SPETVAVAFRTRDRSGNNAKWCSIOHLPVYLPKPSVENEYTLVTERALDREERTYNI 419  
 DB 361 AEETVAAVFSVDLDSGNGKISCSIOEDLPFLKSSGNGEYTLTLTERALDREERTYNI 420

QY 420 TTVTDLGPRKLTQHNLTVTVSVYNDNAPFESOTTYLRYRENNSPALHIGVSATRD 479  
 DB 421 TTVTDLGPRKLTQHNLTVTVSVYNDNAPFESOTTYLRYRENNSPALHIGVSATRD 480

QY 480 SCANQAVTSLPPLPDPPLGSLVSTINADNGOLFALSLDLEALCAEFEEVGAADRSP 539  
 DB 481 SGTNAQVTVSLPPLPDPPLGSLVSTINADNGOLFALSLDLEALCAEFEEVGAADRSP 540

QY 540 ALSSALVAVLVADANDNAPVLYPLONGSAPCTELVPRAAEAGYLVAKVAVDSDGON 599  
 DB 541 ALSSALVAVLVADANDNAPVLYPLONGSAPCTELVPRAAEAGYLVAKVAVDSDGON 600

QY 600 AMLSTQLKATPEPLGEGVMAHNGEVRTA 627  
 DB 601 AMLSTQLKATPEPLGEGVMAHNGEVRTA 628

RESULT 15  
 CDBD\_HUMAN  
 ID CDBD\_HUMAN STANDARD: PRT: 798 AA.  
 AC Q9YSE0:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protocadherin beta 13 precursor (PCDH-beta13).  
 GN PCDHB13.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99308636; PubMed=10380929;  
 RA Wu Q., Maniatis T.,  
 RT "A striking organization of a large family of human neural cadherin-  
 RL like cell adhesion genes";  
 RN Cell 97:779-790(1999).  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21223055; PubMed=11322959;  
 RA Vanhaest K., Kools P., Vanden Eynde E., van Roy F.;

RT "The human and murine protocadherin-beta one-exon gene families show  
 RT high evolutionary conservation, despite the difference in gene  
 RT number.";  
 RL FEBS Lett. 495:120-125(2001).  
 CC - FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.  
 CC - SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC - SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>  
 CC or send an email to [license@isb-stb.ch](mailto:license@isb-stb.ch)).  
 CC -----  
 DR EMBL: AF152492; AAD43753.1; -;  
 DR EMBL: AF217745; AAK51613.1; -;  
 DR HSSP: P15116; INCU.  
 DR Genew: HGNC:8684; PCDHB13.  
 DR MIM: 606339; -;  
 DR MIM: 604967; -;  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin; 5.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART: SM00112; CA; 6.  
 DR PROSITE: PS00232; CADHERIN\_1; 5.  
 DR PROSITE: PS00268; CADHERIN\_2; 5.  
 DR KEGG: K04401; Cell adhesion; Glycoprotein; Signal; Repeat;  
 KW Transmembrane; Multigene family.  
 FT SIGNAL 1 28  
 FT CHAIN 29 798  
 FT DOMAIN 29 690  
 FT TRANSMEM 712 798  
 FT DOMAIN 691 711  
 FT DOMAIN 712 798  
 FT DOMAIN 36 134  
 FT DOMAIN 139 243  
 FT DOMAIN 248 348  
 FT DOMAIN 353 451  
 FT DOMAIN 456 561  
 FT DOMAIN 568 671  
 FT CARBOHYD 418 436  
 FT CARBOHYD 436 456  
 FT CARBOHYD 567 587  
 SQ SEQUENCE 798 AA: 87551 MW: 24478554.498B41A3 CRG64;

Query Match 66.8%; Score 2142; DB 1; Length 798;  
 Best Local Similarity 68.0%; Pred. No. 2e-135;  
 Matches 427; Conservative 66; Mismatches 133; Indels 2; Gaps 2;

QY 1 MEARVVALOKROVSLVFLGVSMAG-AEPLKRYFAEETERTFLANLAIDGLGVEEL 59  
 DB 1 MEASGLICROROVLSFLLGLSLAGAAEPKRSYVEETEGSSFTVNLADLGLEOREF 60

QY 60 SARCRIVSDETIGFLLNPLTGLDLNKLDEEELGPEPCVLPFOLLKPEFQIFRA 119  
 DB 61 SRGRVAVSRGKHLQNLQETADLLNEKLDKEDLCGHTPCVLRVYLLSEPEFQIA 120

QY 120 ELWVDINDHSPVFLDREITLNIESTTPGATFLLSAHSDVGINLNRNYTSSNVYFH 179  
 DB 121 ELQVIDINDHSPVFLDKQMLVKSSESPGTAFLKNAEDDIGNNIENIISNSYFR 180

QY 180 INVHNDGEGNVSELYVDKVDREVPRLTLTGDLGGSPRSGTTLIRLVLDINDV 239  
 DB 181 VLTFRKSDGRKPELVLDKLDREBEALRLTLALDGGSPRSGTAQVYIEVDVNDNA 240

QY 240 PEFEVSLKYQVENSPPVGLVTVSARDLDTGSGNEIYAFVAFERTLKTFRINSNG 299  
 DB 241 PEFGQPFYRVQISEDSPISEFLVYKSAIDVDVTGNGEISTSLFQASDEISTKFKVDLTG 300

QY 300 NLHLKAEINTEAIOYTLTLTQAKDGGSLGKCTVVVHTDINDPELLMSLTSPIEN 359  
 DB 301 EIRLKQDLEKQSYEVNIEARDAAGSGKCTVLIQYIDVNDHAFVETMAFSPIDEN 360

OY 360 SPETVAVFRINDRDSGNNAKWCSTIOHLPFVLKPSVENETTLVTERALDREERTENI 419  
:|||||: | |||: |||: |||: |||: |||: |||: |||: |||  
Db 361 APEYVALFVSVDLDGSENGKISCSIOEDLPFLK-SAEFYTLTERPLDRESRAEYNI 419  
OY 420 TITVTDIGTPRLKTOHNLTVSDVNDNAPTESOTTITRVRENNSPALHIGSVSATDRD 479  
|||||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 420 TITVTDIGTPRLKTOHNLTVSDVNDNAPTESOTTITRVRENNSPALHIGSVSATDRD 479  
OY 480 SGANQVYYSLLPDPOLPLGSLVSINADNGOLFALRSIDFEALQAFEFYRGADRGSP 539  
|||||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 480 SGTNQVYYSLLPDPOLPLGSLVSINADNGOLFALRSIDFEALQAFEFYRGADRGSP 539  
OY 540 ALSSQALYRVLVADANDNAPFVLPLONGSAPCTELVPRAAEGYLVAKVAVDGDGON 599  
|||||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 540 ALSSQALYRVLVADANDNAPFVLPLONGSAPCTELVPRAAEGYLVAKVAVDGDGON 599  
OY 600 AMLSTQLLKATEPGLGFWAHNGEVRTA 627  
|||||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 600 AMLSTQLLKATEPGLGFWAHNGEVRTA 627

Search completed: May 16, 2003, 10:46:24  
Job time : 19 secs